July 13, 2010

FEDERAL COMMUNICATIONS COMMISSION
Office of the Secretary

From:
Professor Scott Jordan
Department of Computer Science
University of California, Irvine
3029 Bren Hall
Irvine, CA 92697
Email: sjordan@uci.edu
Phone: 949-824-2177
Webpage: http://www.ics.uci.edu/~sjordan

Re: Comments Regarding the Notice of Inquiry in the Matter of a Framework for Broadband Internet
Service (GN Docket No. 10-127)

Scott Jordan is a **Professor of Computer Science** at the University of California, Irvine. He is
participating in the **Technical Advisory Process** as part of the Open Internet proceeding. His research
focuses on traffic management, including both technical mechanisms and public policy.

The comments here do not necessarily represent the views of anyone but him.

The following comments are presented in response to the following two questions in the NOI:

> *60. Some have suggested that the Commission should take account of the different network "layers" that compose the Internet. Are distinctions between the functional "layers" that compose the Internet relevant and useful for classifying broadband Internet service? For example, the Commission could distinguish between physical, logical, and content and application layers, and identify some of those layers as elements of a telecommunications service and others as elements of an information service. (As discussed above, the Commission historically has distinguished between Internet connectivity functions and Internet applications.) If the Commission adopted this approach, which of the services offered by wired broadband Internet service providers should be included in each category? Are the boundaries of each layer sufficiently clear that such an approach would be workable in practice? Would such an approach have implications for services other than broadband Internet service?*

> *63. If the Commission were to classify a service provided as part of the broadband Internet service bundle as a telecommunications service, it would be necessary to define what is being so classified. Here we ask commenters to propose approaches to defining the telecommunications service offered as part of wired broadband Internet service, assuming that the Commission finds a separate telecommunications service is being offered today, or must be offered.*

(1) The important layer boundary lies between the network and transport layers, not between the physical and logical layers or between the logical and content layers.

The network layers and the layers below it (commonly known as layers 1-3 in either the OSI or Internet layered models) display quite different technical and economic characteristics than the transport layer and the layers above it (commonly known as layers 4-7 in either the OSI or Internet layered models).

From a technical point of view, the network layer and the layers below it must be implemented **in every portion of the network**, including the access network, while the transport layer and the layers above it are traditionally implemented **only at the network endpoints**. As a consequence, the network layer and the layers below it support network functions that must be implemented in every portion of the network, including routing, addressing, and Quality-of-Service (QoS). In contrast, the transport layer and the layers above it support network functions that are implemented at clients, servers, and peers, including email, web-hosting, and caching.

From an economic point of view, the network layer and the layers below it exhibit a **high barrier-to-entry**, resulting in a small number of carriers offering service in any particular location. In contrast, the transport layer and the layers above it exhibit a **low barrier-to-entry**, resulting in a competitive market with a large number of application providers.

This layered delineation also helps with correct application of the **end-to-end principle**. The end-to-end principle suggests that network functionality should be implemented in network infrastructure only if it cannot be implemented effectively in end devices or applications. With the recognition that network layer and the layers below it are implemented in every hop along the transmission path, whereas the transport layer and the layers above it are only implemented in end devices, the end-to-end principle suggests that network functionality should be implemented

at or below the network layer only if it cannot be implemented effectively at and above the transport layer.

This delineation agrees with the intent of the past use of the terms "telecommunication services" and "information services". The network layer and the layers below it do not change the content of information, similar to telecommunication services. In contrast, the transport layer and the layers above it create, store, or change the presentation of information, similar to information services.

The boundary between the network and transport layers is well recognized by a host of **Internet standards**. The approach would thus be workable in practice.

I recommend that the telecommunications component of broadband Internet service be defined as those services **"that reside at or below the network layer or are required to manage the network and that do not fall under Title VI of the Communications Act"**. The inclusion of services required to manage the network is designed to include a small number of services that reside at or above the transport layer, that are required for Internet access, and that are currently not offered competitively, e.g. IP address assignment using Dynamic Host Control Protocol (DHCP). The exclusion of services that fall under Title VI of the Communications Act is designed to allow Internet Service Providers to have a choice between following the requirements of Open Internet access and following the requirements of Title VI solely for those services that are offered entirely within their own network.

While Internet architecture is based on this distinction between the network and the layers below it versus the transport layer and the layers above it, the communications policy literature and the FCC have not consistently placed the boundary between telecommunication services and information services in the correct place. Much of the literature discusses a "**logical layer**" which often combines the network and transport layers (as well as other layers). However, Internet architecture does not recognize a logical layer. Furthermore, the network and transport layers display quite different technical and economic characteristics, and the network layer should be classified as a telecommunications service while the transport layer should be classified as an information service.

(2) Network layers are a simple and effective way to classify the telecommunications component of broadband Internet service and to implement the goals of an Open Internet.

Internet Service Providers (ISPs) should be ensured that they have the right to apply network management mechanisms that do not threaten a level playing field, including the use of QoS. However, ISPs should be prohibited from refusing to provide, or providing at inflated prices, telecommunications service to competing providers of information services in order to differentiate or favor the ISP's own application offerings. ISPs should also be prohibited from making exclusive deals to provide telecommunications service to certain application providers. ISPs should also be ensured that regulation does not impede an ISP from making arrangements with consumers, application providers, and peering ISPs for telecommunications service in a manner that does not conflict with the above goals.

This balancing act between allowing reasonable network management while simultaneously prohibiting use of telecommunications service that threatens a level playing field can be **simply and effectively implemented by using network layers to define the telecommunications**

component of broadband Internet service. The prohibitions should apply only to those services that must be implemented in an ISP's network in order for the application to work effectively. Services that reside at or below the network layer or required to manage the network fall into this category, and thus should be classified as telecommunication services. Services that reside at or above the transport layer and are not required to manage the network do not fall into this category, and thus should be classified as information services.

In particular with respect to QoS, an ISP should be ensured that it can implement QoS at any network layer. However, if an ISP implements QoS at or below the network layer (as one would expect), then it should be prohibited from using this service to tilt the playing field of applications offered at or above the transport layer.

Reasonable traffic management practices that are implemented at or above the transport layer, e.g. caching of content, can be implemented not only by a broadband Internet service provider but also by competing non-facilities based providers. Such higher layer services should be classified as information services and should be excluded from Open Internet regulations.

The FCC, however, should also retain the authority to **prohibit unreasonable traffic management practices implemented at or above the transport layer**. I have recommended that a traffic management practice implemented in transit routers be considered unreasonable if it blocks connections or prioritizes traffic based on source, destination, service provider, or unreasonable application provider payment, and is not done by user choice.

With this layered approach to defining telecommunication service, there is no need for an exception for "**managed services**". The intent of creating a category of "managed services" seems to be to allow implementation of QoS to enable high quality services. However, if managed services are not subject to the same rules as unmanaged services, then providers of broadband Internet access service **may be allowed to implement QoS solely to support their own managed services**. This would completely undermine the Commission's goal of promoting competition in services that require QoS. **The need for a separate class of managed services can be eliminated by refining the rule on nondiscrimination to apply only to telecommunications services**, as defined above. With this refinement, broadband Internet access providers would be allowed to implement QoS at or below the network layer to support voice and video services without being required to designate such services as managed services. However, if a broadband Internet access service provider uses QoS to support its own services, then it would be required to make available QoS at reasonable prices to its own subscribers and to the provider's peering providers.

(3) The same approach works effectively for other types of networks, including wireless networks.

This same layered approach can be simply and effectively applied to other types of networks, including wireless networks and telephone networks.

Although wireless networks require substantially greater traffic management than wired networks, these differences occur only below the transport layer. Therefore, the same rules should apply to wireless and wired networks, and the application of the rules to wireless and wired networks **should only differ below the transport layer**.

A proper layered definition of telecommunication services can address this concern effectively, since the rules can be equally applied to all forms of broadband Internet access, independent of the technology used. Wireless networks may use more stringent traffic practices below the

transport layer, but these would not violate any of the proposed rules. Wireless networks, as with wired networks, should allow access to lower layer QoS through an interface at the network layer. Such an interface would then allow wireless broadband Internet access service providers to practice efficient and effective traffic management, without the need for them to prohibit particular applications or prohibit tethering.

Taking a longer term view, for at least 25 years there has been a trend toward technology convergence of the Internet, telephone networks, cable networks, and wireless networks. Although lower layer protocols will continue to differ based on the characteristics of the physical connection, these four types of networks will continue to converge at the network layer and above, and to offer an increasingly similar set of highly diversified services and applications. In due course, I hope that Congress will recognize this technology convergence in a thorough rewrite of the Communications Act, and I hope that this rewrite will similarly recognize the importance of this interface between the network and transport layers.

(4) A layered approach can be used in conjunction with Section 202 to ensure an Open Internet and to define reasonable network management, and in conjunction with Section 254 to apply universal service to converged networks.

The following attached papers give more detail:

Scott Jordan, *A Layered Network Approach to Net Neutrality*, International Journal of Communication, vol. 1, 2007, pp. 427-460.

Scott Jordan, *Do Wireless Networks Merit Different Net Neutrality than Wired Networks?*, Research Conference on Communication, Information and Internet Policy (TPRC), 2010.

Scott Jordan and Arijit Ghosh, *How to Determine Whether a Traffic Management Practice is Reasonable*, Research Conference on Communication, Information and Internet Policy (TPRC), 2009.

Scott Jordan, *A Layered United States Universal Service Fund for an Everything-over-IP world*, Telecommunications Policy, vol. 33, 2009, pp. 111-128.

Respectfully submitted,

Scott Jordan
Professor of Computer Science

net neutrality issue

A Layered Network Approach to Net Neutrality[1]

SCOTT JORDAN

University of California, Irvine

Net neutrality was the most contentious communications policy issue considered by Congress during 2005-2006. The issue is the result of a fragmented communications policy unable to deal with technology convergence. In this paper, we develop a net neutrality policy based on the layered structure of the Internet that gracefully accommodates convergence. Our framework distinguishes between discrimination in high barrier-to-entry network infrastructure and in low barrier-to-entry applications. The policy prohibits use of Internet infrastructure to produce an uneven playing field in Internet applications. In this manner, the policy restricts an Internet service provider's ability to discriminate in a manner that extracts oligopoly rents, while simultaneously ensuring that ISPs can use desirable forms of network management. We illustrate how this net neutrality policy can draw upon current communications law through draft statute language. We believe this approach is well grounded in both technology and policy, and that it illustrates a middle ground that may even be somewhat agreeable to the opposing forces on this issue.

1.   Introduction

During 2005 and 2006, Congress worked on the most substantial rewrite of the nation's communications law in a decade. The issue initially motivating this task in both houses of Congress was video franchising, which attempted to streamline the laws regulating carriers offering broadcast video services, with the goal of enabling faster deployment of video products by telephone companies. The House bill included not only video franchising, but also a weak version of net neutrality, voice over IP (VoIP) interconnection, e911, and municipal broadband. The Senate bill included an even wider range of

---

[1] These ideas do not necessarily represent the views of the University of California, IEEE-USA, or Senator Ben Nelson. Portions of this work also appear in Jordan (2007). Whereas Jordan (2007) is directed at the networking community and focuses on the relationship of policy to Internet architecture, this paper is directed at the policy community and focuses on development of statute language.

Scott Jordan: **sjordan@uci.edu**
Date submitted: 2007-06-07

issues, including all of the issues in the House bill, plus universal service, broadcast flag, white spaces, interoperable emergency communications, DTV, wireless preemption, tax moratoriums, and many other issues. Among all of these communication issues, net neutrality was easily the most contentious. Lack of resolution of net neutrality was generally given credit for derailing the communications bill.

Net neutrality represents the idea that Internet users are entitled to service that does not discriminate on the basis of source, destination, or ownership of Internet traffic. The idea is rooted in the manner in which the Internet has historically operated, in which all traffic is forwarded as quickly as possible, with limited differentiation based on the application and without any performance guarantees. There is great disagreement, however, about the future implications of this relatively simple idea as the Internet progresses and as the economic communications landscape changes.

Proponents of net neutrality (generally, application providers and consumer groups) argue that without a prohibition on discrimination, Internet Service Providers[2] (ISPs) may charge application providers discriminatory prices for access to dedicated bandwidth or for quality of service (QoS), or may outright block access to certain applications or websites, and that such activity will inhibit development of new Internet applications. Most proponents believe that ISPs should not be allowed to charge for priority access to the Internet portion of their service offerings.

Opponents of net neutrality (generally, ISPs) argue that there is no current problem, that competition is sufficient to ensure that commercially negotiated arrangements for bandwidth or QoS will not negatively impact consumers, and that any regulation will discourage investment in network infrastructure.

The topic is timely for a combination of technological and policy reasons. On the technology side, there is a steady convergence between the networks traditionally used to offer voice, video, and data services. Most networking technologists expect that the technical differences between telephone networks, video networks, cellular networks, and the Internet to diminish in future years, with all of these networks becoming capable of efficiently supporting a combination of voice, video, and data services. Competition between the carriers supporting the infrastructure of these various networks and application providers offering a wide range of services will thus intensify.

On the policy side, U.S. federal communications policy was separately developed for telephone networks, cable video networks, and cellular networks. In the absence of any explicit statutes regarding Internet access or services, the Federal Communications Commission (FCC) recently declared that Internet access is not subject to the same common carrier regulation that addresses telephone networks,

---

[2] In this paper, we use the term ISP to represent Internet access providers, namely carriers who provide last mile access to residential and business customers. ISPs can provide two types of service - access to Internet infrastructure, e.g., DSL, and application services, e.g., e-mail. This distinction is made later in the paper when we introduce definitions to be used in statute language.

which had the effect of removing from Internet access several prohibitions on discrimination that were included in common carrier regulation. This lifting of discrimination constraints triggered the push for net neutrality.

Net neutrality was a heavily lobbied issue in 2006. Estimates of the amount spent on net neutrality advertising and lobbying range up to $100 million[3]. The issue also attracted wide attention on the Internet and in the media. The term *net neutrality* appears on more than two million web pages[4]. Net neutrality was the topic of editorials, commentaries, and news articles in dozens of national and local newspapers and magazines, and garnered some attention on television, ranging from *Moyers On America* to *The Daily Show*.

The academic literature on net neutrality has been split on the issue. Openists (Bar et al., 2000; Lemley & Lessig 2001) believe that the Internet is best served by maintaining a *dumb network* that does not differentiate among different types of traffic. They support a policy based on *open access*, in which Internet infrastructure and applications can not be bundled using either technical or business mechanisms. They argue that vertical integration harms consumers, that most innovation comes from application providers, and that open access will maximize social welfare. Openists support a strong version of network neutrality.

Deregulationists[5] (Owen & Rosston 2003; Yoo 2005) believe that ISPs are in the best position to determine the most beneficial evolution of the Internet. They expect that the Internet will become a *smart network* that uses traffic discrimination in order to accomplish product differentiation. They support a policy that deregulates the Internet in a manner that allows ISPs to vertically integrate, bundle services, and use traffic discrimination as they see fit. They argue regulation will hinder investment by ISPs and that in the absence of regulation ISPs will only vertically integrate in ways that maximize consumer welfare. Deregulationists oppose network neutrality.

There is a limited amount of academic literature that attempts to strike middle ground. Nondiscriminationists (Wu 2004; Peha 2006) believe that there are good and bad uses of traffic discrimination. They support a policy that allows vertical integration and traffic differentiation, but restricts their use to ensure that ISPs do not discriminate in a manner that extracts oligopoly rents. They argue that such a balanced approach will allow development of a smart network in a manner that does not restrict development of applications. They would support a limited version of net neutrality that falls short of open access, but have not yet fully formulated such a policy.

Finally, there is an academic literature that addresses net neutrality and related topics in the context of Internet architecture (Clark, Wroclawski, Sollins, & Braden 2005), broader revisions of

---

[3] See e.g., Bloomberg (2006).

[4] Google search on 6/14/2007.

[5] The terms *openist* and *deregulationist* were introduced in Wu (2004).

communications policy (Bar & Sandvig 2000; Werbach 2002; Solum & Chung 2003), economic analysis (Speta 2000; Farrell & Weiser 2003; van Schewick 2007), and market analysis (Lehr, Gillett, Sirbu, & Peha 2006).

In this paper, we argue that net neutrality is, at its core, an attempt to address problems posed by a fragmented communications policy unable to deal with technology convergence. We adopt an approach jointly grounded in Internet technology and communications policy. We argue that the evolving layered Internet architecture supports the model of a smart Internet that allows only certain types of discrimination. We accept the premise that vertical integration between infrastructure and applications poses potential threats to a level playing field. We suggest that an important tool in solving such problems is a proper delineation of Internet infrastructure and Internet applications. We illustrate how such a delineation can be used to restrict an ISPs ability to extract oligopoly rents through discrimination, while simultaneously ensuring that ISPs can use desirable forms of network management. We further illustrate how this use of layering can appropriately limit the scope of regulation. Finally, we suggest that net neutrality can be addressed in a manner consistent with current Federal communications law, and we propose draft statute language on this basis.

Our analysis is heavily based on communications technology and historical communications policy. Our approach belongs in the Nondiscriminationist camp. Our solution essentially mandates the use of open interfaces, but does not mandate full open access.[6] Although we believe that there are many other interesting ways to address net neutrality, we hope that our convergence-inspired layered approach will illustrate a solution that is well grounded in both technology and policy, and that may even be somewhat agreeable to many of the opposing forces on this issue.

The paper proceeds as follows. In section 2, we briefly discuss technology convergence, explain the layered Internet architecture model, and analyze evolving technological developments. In section 3, we review the relevant academic literature, present pro and anti net neutrality lobby arguments, and characterize the various congressional approaches to net neutrality. In section 4, we illustrate how a proper delineation of Internet infrastructure and Internet applications can serve as a foundation for a new net neutrality policy that restricts an ISP's ability to discriminate while allowing reasonable network management. In section 5, we delineate acceptable and unacceptable uses of network management. Finally, in section 6, we look to current Federal communications law for guidance, and translate our policy into draft statute language.

2.   Communications Technology

2.1   Convergence

In operation since the 1890s, telephone networks have traditionally offered voice service. In the past 25 years, most telephone companies have deployed fiber in portions of their networks, and transitioned from analog to digital transmission in the network backbone. With the extra capacity and

---

[6] We do not take a position here on open access, as this is outside the scope of the paper.

capabilities this provides, most telephone companies now offer voice service, video conferencing, and Internet access. Most of the Internet backbone runs over telephone networks. In addition, many telephone companies have recently announced plans to offer video service.

Cable video networks were the next significant entry into the field, and were designed to offer broadcast video service. Since their introduction in the 1970s, subscribership to cable and DBS networks has increased to about 86% of U.S. households. Earlier networks were analog and broadcast-only systems. Recently, many cable systems have migrated to digital transmission, which allow for a vastly increased number of channels, Internet access, and telephone service.

Cellular networks were introduced in the 1980s, and were initially designed to offer voice service. The first generation of cellular phones were built on an analog platform, but the second and third generations are digital. In the U.S., there is now an average of about 1.7 cell phones per household, and the number of cellular phones has recently eclipsed the number of wired phones. Future wireless networks are expected to transition to an architecture that more closely resembles the Internet, and wireless data services are expected to grow quickly.

The Internet, created in 1969 by the Department of Defense Advanced Research Projects Agency, was designed to offer file transfer service. Taking advantage of openly published rules of operation and freely distributed software, many research and educational institutions attached their computers to the Internet during the 1970s. The network has largely blossomed, however, due to the introduction of personal computers during the 1980s and the development of the World Wide Web in the 1990s. In the U.S., approximately 58% of households have Internet access[7]. The Internet is now in the process of transitioning toward an architecture that can more efficiently support real-time applications such as voice and video.

These four technologies (telephone networks, cable video networks, cellular networks, and the Internet) began to merge in the 1980s, with the introduction of optical fiber and advanced computer capabilities. Economic and regulatory forces have reinforced the technical trends. Recent trends are based on an increased desire for multimedia involving voice, data, and video. A flurry of mergers between communications companies has resulted, bringing together content providers, broadcasters, cable TV networks, local and long distance telephone networks, wireless networks, and Internet service providers.

Most networking technologists expect that the technical differences between telephone networks, video networks, cellular networks, and the Internet to diminish in future years, with all of these networks becoming capable of efficiently supporting a combination of voice, video, and data services.

Convergence is a principal cause for net neutrality's timeliness, since the motivation for differentiation between different Internet traffic lies in the carriage of voice and video traffic. We turn in section 2.3 to these specific developments. First, however, we must discuss layered Internet architecture.

---

[7] Approximately 73% of households that have Internet access use broadband.

2.2   Layered Internet architecture

Telephone networks, cable video networks, cellular networks, and the Internet are all based on the concept of a layered architecture. Each network device, and the network as a whole, is abstractly modeled as being composed of a number of vertical layers. Each layer provides certain functionalities.

Layering is a form of modularity. In modular design, a designer of a module need only understand the functionality and the *interface*, not the detailed operation, of other interoperating modules. Although designing a component in a modular fashion restricts the design space, the benefits usually outweigh the costs.

In a layered architecture, it is useful to think of two types of interoperation. First, within a network device, modules are only allowed to interoperate with other modules that are at most one layer above or below; typically, the higher layer requests the lower layer to accomplish some task, and the lower layer eventually responds with a message indicating success or failure. Second, modules that are part of two separate network devices are only allowed to interoperate with each other if they are members of the same layer. The benefit of these rules is that a very large number of network devices using a large number of network protocols can interoperate with each other. For instance, you and I can send each other e-mail even though we may use different e-mail programs.

The reference model for layered architectures is the OSI model, developed by the International Standards Organization. The OSI model is composed of 7 layers, as pictured in figure 1. It is useful to think of the physical connection, e.g., wire, as being located below the bottom-most layer (layer 1) and the user, e.g., you, as being located above the top-most layer (layer 7).

OSI layers		Internet layers
7: applications & content	application	application
6: data formatting	presentation	
5: session management	session	
4: end-to-end connection	transport	transport (TCP)
3: routing	network	network (IP)
2: packet to bits	data link	LAN-link
1: bit to analog signal	physical	

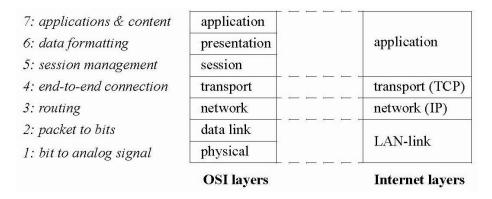**OSI layers**   **Internet layers**

Figure 1: OSI and Internet layered models

OSI layer 1, called the *physical layer*, implements encodes a bit into a physical signal and vice versa. OSI layer 2, the *data link layer*, translates a packet into a set of bits and vice versa, and implements a set of rules (called a *protocol*) about which device can transmit when. OSI layer 3, the *network layer*, is concerned with routing a packet from one network device to the next. OSI layer 4, the *transport layer*, is concerned with functionality required to form a complete connection between a source and destination, including dealing with lost packets and responding to congestion. OSI layer 5, the *session layer*, manages an entire communication session, e.g., logging onto a service. OSI layers 6, the *presentation layer*, concerns data presentation, e.g., file or video compression. Finally, OSI layer 7, the *application layer*, deals with user applications and other high-level functionality, e.g., web browsing, e-mail, file transfer, file sharing, instant messaging, gaming, etc.

Not every network device contains all 7 layers. Personal computers do contain all 7 layers – the network interface card (e.g., Ethernet card) implements OSI layers 1 and 2, the operating system (e.g., Windows) implements OSI layer 3 and part of layers 4 through 7, and user-installed software implements the remainder of layers 4 through 7. A network router, however, often contains only layers 1 through 3. As a result, communication from source to destination follows a complicated path as pictured in figure 2.
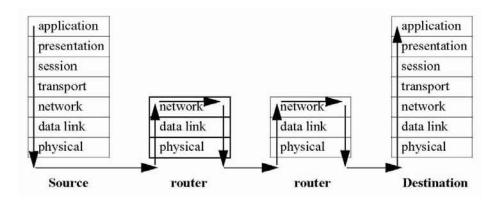

Figure 2: Routing

As an example, consider an e-mail sent from me to you. The e-mail is generated through my e-mail program on my computer (the source) at layer 7 (top left). Layer 7 hands the e-mail off to layer 6, which deals with encoding of the e-mail, e.g., html. Layer 6 then hands the e-mail off to layer 5, which logs onto my mailserver[8]. Layer 4 breaks the e-mail into a sequence of packets, and hands them one by one to layer 3. Layer 3 determines the first router on the path from me to you. Layer 2 decides when each packet can be transmitted, and layer 1 transmits each packet bit by bit. When these bits are received by the first router, layer 1 on that router translates the received signal into bits and layer 2 puts the bits back together into a packet, allowing layer 3 to look at the packet and determine the next router along the path. The second router acts similarly to the first router. Finally, layers 1 through 4 on your computer (the destination) assemble the received signal back together into the e-mail, layer 5 retrieves the e-mail from your mailserver, and layers 6 and 7 display the e-mail in your e-mail program.

[8] For simplicity, we assume here that the mailservers reside on the source and destination computers.

Layers 1 through 3 can be thought of as network infrastructure, while layers 4 through 7 can be thought of as network applications. This distinction will be discussed in much more detail below as we consider the impact of layered architectures upon communications policy.

Although the OSI model serves as a reference for all network architectures, different networks have modified the model for their own use. As an example, the Internet uses a model with a reduced number of layers, as pictured in figure 1. OSI layers 1 and 2 are combined into a single Internet *LAN-link layer*. OSI layer 3 is also called the Internet *network layer*; it includes the *Internet Protocol* (IP). OSI layer 4 is also called the Internet *transport layer*; it includes the *Transmission Control Protocol* (TCP). OSI layers 5 through 7 are merged into a single Internet *applications layer*.

2.3   Technological developments

We turn now to a brief discussion of technological developments, due to convergence, that underlie the net neutrality issue. We start by considering the requirements of different types of applications. Applications differ in terms of the time scale on which users interact with the application. In real-time applications such as telephone calls or video conferencing, the information sent through the network must be received within a few tenths of a second after it is sent, or the perceived performance will be poor. In non-interactive applications such as e-mail, a delay of tens of seconds is usually still considered good performance. There are also applications with intermediate levels of interactivity, such as web browsing, in which a delay of a few seconds is usually acceptable. Applications also differ in terms of how they define good performance. Voice and video applications tend to base performance on *loss* (percent of packets never received) and *delay* (seconds from source to destination), while less interactive applications tend to base performance on *throughput* (bits per second received).

Traditional Internet transport uses *best-effort service*, in which the network attempts to transmit each packet from source to destination as quickly as possible. All users and all applications sharing a network link share the capacity of that link on a first-come first-served basis in the Internet network layer, without regard to the source or destination of the traffic and without regard to the type of application the packet supports[9]. The result is that congestion on a link degrades the performance of all users and all applications that pass through that link.

Best-effort service is often fine for applications whose perceived performance degrades slowly with increased congestion, e.g., e-mail, file transfer, and web browsing. In contrast, real-time applications such as telephone calls and video conferencing require a particular performance level (or better) to be usable; for these applications, best-effort service is often not sufficient to maintain acceptable performance.

---

[9] Technically, some applications are subject to congestion control measures in the Internet transport layer. However, these measures do not discriminate based on source, destination or application type.

In response, there have been recent technological developments in the Internet which provide differentiated service to different Internet traffic. One type of such differentiated service is to place a limit on the bandwidth used by certain Internet traffic. This approach, often called *traffic shaping*, is appropriate when there is a class of Internet traffic that tends to consume relatively large amounts of capacity, is viewed as low priority to the organization, and whose performance degrades smoothly with congestion.

Another type of such differentiated service is to label Internet traffic with distinct *priority* levels. At each router, in the Internet network layer, traffic with higher priority levels experiences better performance than traffic with lower priority levels. From the user's view, this priority approach is roughly similar to USPS Priority Mail service in that the packages go through the same system as First Class mail but experience lower loss and delay.

Finally, a third type of such differentiated service is to *reserve* network capacity for certain traffic classes and to limit the traffic in these classes. At each router, in the Internet network layer, such selected traffic shares a reserved portion of the network capacity; since this capacity is actively managed, the traffic experiences guaranteed acceptable performance. From the user's perspective, this reservation approach is roughly similar to a toll or carpool lane[10].

The priority and reservation approaches are generically known as *Quality of Service* (QoS) mechanisms. Both QoS approaches give the selected traffic enhanced performance. In contrast, traffic shaping can be viewed as intentional *degradation* of the selected traffic class. It is worth noting that these QoS mechanisms are implemented at the Internet network layer; the consequence of this approach is that QoS must be provided at *every router* along the path from source to destination[11] in order to be effective.

In both traffic shaping and QoS, the decision of which traffic to enhance or degrade can be based on (1) the type of application, (2) the source, (3) the destination, (4) consumer payment, or (5) application provider payment. For example, if priority service is given to all voice and video traffic, or if traffic shaping is applied to file sharing traffic, then application type is the basis for the decision. If capacity is reserved for all voice and video traffic to/from consumers for a fee, then consumer payment is the basis. If capacity is reserved for all traffic to/from application providers for a fee, then application provider payment is the basis.

As traffic shaping and QoS mechanisms have become available in network equipment, many ISPs have adopted their use for various purposes. Some ISPs currently use QoS to guarantee acceptable quality for their own VoIP service[12]. Many universities use traffic shaping to limit file-sharing traffic to a

---

[10] Technically, it is similar to a toll lane in which the traffic is managed so that congestion never occurs in the toll lane.

[11] or at least at every congested router.

[12] See e.g., Cox (2004).

small proportion of the organization's network capacity in order to protect the performance of other traffic viewed as more important to the organization's mission.

Use of these mechanisms is expected to grow with technology convergence. Most of the large carriers have announced plans to deploy QoS mechanisms, often in coordination with deployment of fiber and video service. In the near term, priorities or reservations will be used for selected traffic that both originates and terminates within the carriers network, or that transits onto the public switched telephone network. In this case, a carrier's own network management can provide acceptable performance to limited traffic classes using QoS techniques. In the long term, we expect that QoS will also be applied to selected traffic that originates or terminates (but not both) within the ISP's network and transits onto another carrier's portion of the Internet. In this case, acceptable performance may only be provided through cooperation with other carriers offering QoS.

We expect the initial use of QoS will be to support a carrier's own VoIP and video services. It is unclear whether carriers will offer QoS to competitors' applications, on the basis of application type, consumer payment, and/or application provider payment.

We believe these developments are at the core of the issue of net neutrality. Specifically, we are concerned about the basis on which differentiated service may be used to support real-time applications and applications that require relatively large amounts of bandwidth. In contrast, we are *not* concerned about applications that are less interactive and require little bandwidth, such as most web-browsing; we do not foresee best-effort Internet transport degrading to the point where such applications do not receive acceptable performance.

In addition, as networks converge, it is becoming less clear what portion of the integrated network is considered to be the public Internet and what portion is considered to be a private network. Specifically, in the case of traffic that does not transit onto another carriers portion of the Internet, the distinction between Internet traffic, VoIP traffic, and video traffic may be more a matter of user perception than of technical distinction. This causes a fundamental problem with separate regulation Internet traffic, telephone traffic, and video traffic.

3.   Arguments For and Against Net Neutrality

3.1   Academic Literature

The academic literature on net neutrality is rooted in an earlier literature on open access, which in turn partially relies on an older literature on Internet architecture.

As discussed in section 2.2, the Internet is based on a layered architecture. As pictured in figure 2, the full 7 OSI layers only exist at the source and the destination (the endpoints), not at the

intermediate routers. As a result, functions in OSI layers 1 through 3[13], such as routing, must be implemented in each router, as well as the source and destination computer. However, functions in OSI layers 4 through 7[14], such as e-mail upload and download, are only implemented in the source and destination computers, not in each router.

A closely related design paradigm is called the *end-to-end principle* (Saltzer, Reed, & Clark, 1984). The principle suggests that network functionality should be implemented in OSI layers 1 through 3, and hence in each router, only if it cannot be implemented effectively in higher layers. This principle has been followed in much (but not all) of Internet design to date. As a consequence, although there are a large number of protocols at the Internet LAN-link layer to accommodate transmission over various media (e.g., Ethernet, WiFi, DSL, cable modem, CDMA), and there are a very large number of protocols at the Internet application layer to support various applications (e.g., pop, smtp, http, p2p, IM), there are relatively few protocols at the intervening Internet network and transport layers (principally IP, TCP, and UDP).

The Openists believe that the end-to-end principle is responsible for the tremendous amount of innovation at the application layer. They argue that communications policy should mandate open access to the network infrastructure for application providers in order to assure continued innovation. The debate over open Internet access was precipitated by FCC consideration of whether to allow ISPs[15] to bundle Internet access with applications such as e-mail and web hosting[16]. Bar et al. (2000) suggest that policy for Internet access should be developed on the model of open access mandated for the telephone network in the *Telecommunications Act of 1996*, by mandating availability on cost-effective terms of key network components to service providers who do not own Internet access infrastructure. Lemley and Lessig (2001) support a policy in which Internet infrastructure and applications can not be bundled using either technical or business mechanisms. They argue that an open network encourages greater innovation than a closed network and that vertical integration harms consumers. They also demonstrate how open access leads to net neutrality, by using the end-to-end principle as an argument for maintaining a dumb network that does not differentiate among different types of Internet traffic. Herman (2007) adds first amendment arguments to support net neutrality.

The Deregulationists believe that there is a trend away from the end-to-end principle. They believe that ISPs are in the best position to determine the most beneficial evolution of the Internet, and that this will likely entail a smart network that uses traffic discrimination in order to accomplish product differentiation. Speta (2000) argues against open Internet access. He postulates that vertical integration

---

[13] or equivalently in the Internet LAN-link or network layers

[14] or equivalently in the Internet transport or application layers

[15] There is an unfortunate inconsistency in the use of the term *ISP*. In this paper, we use the term ISP to mean carriers who provide last mile access to residential and business customers. However, in the open Internet access debate, the term ISP was often used to mean providers of services such as e-mail and web-hosting who do not own Internet access infrastructure.

[16] The FCC later allowed such bundling. We discuss this in more detail below.

will not threaten competition among content providers, on the basis of an economic theory of network externalities that states that monopolists will only vertically integrate in ways that maximize consumer welfare. He argues that vertical integration may be necessary to justify investment in network infrastructure, and that ISPs will be motivated to allow competition to their own applications since this will add value to the broadband connection they sell. Owen and Rosston (2003) view net neutrality as a property rights issue. They claim that it is unknown whether welfare is maximized by assignment of property rights to ISPs or to application providers, and therefore that net neutrality regulation should not be imposed. Yoo (2005) extends the anti open Internet access argument put forward by Speta to net neutrality. He believes that the theory of network externalities undermines the need for mandated net neutrality, and furthermore that the lack of concentration of broadband access when measured at a national level predicts that ISPs will not have market power in the application space. Yoo also suggests that competition can be more effectively fostered between different physical networks that are matched to different types of applications. Yoo (2006) furthermore suggests that exclusive arrangements between ISPs and application providers can be justified on the basis of the efficiency of traffic shaping as a proxy for usage-based pricing. Finally, McTaggert (2006) argues that the Internet is not neutral, using examples of Internet portals, caching, routing, and QoS.

Some economists counter that application of the theory of network externalities to open access or net neutrality is not so straightforward. Farrell and Weiser (2003) discuss potential problems with application of the theory to open access, and suggest that uncertainty dictates further study of the effect of various types of Internet regulation in other countries. Van Schewick (2007) argues that a correct application of the theory demonstrates that ISPs do have an incentive to discriminate against competing application providers, and that the economic benefits of net neutrality outweigh the costs.

Nondiscriminationists see more complexity in the net neutrality issue than the Openists or Deregulationists. Wu (2003) supports net neutrality but opposes open access. He is not opposed to vertical integration, since he believes that QoS may require it. He argues that open access is inferior to direct rules on discrimination that ensure that harmful types of discrimination are not used by ISPs. He suggests that such a policy might be constructed by allowing discrimination at the Internet LAN-link layer but not at the Internet network layer. As a result, he would allow bandwidth tiers and some forms of QoS, but discrimination on the basis of the type of application or on the source or destination would be prohibited. Peha (2006) surveys various technical methods of discrimination, and argues that some uses are beneficial and some are not. He discusses dangers from harmful uses of discrimination, focusing on oligopoly rents in upstream markets. He suggests that good policy might be based on lists of allowed and banned discrimination.

3.2 Lobbyist Arguments

The pro net neutrality lobby falls roughly into two categories - application providers (led by Google, Amazon, Yahoo!, eBay, and Microsoft) and consumer groups (led by moveon.org, Free Press, Consumers Union, and Common Cause).

The application providers are roughly in the Nondiscriminationist camp. They are primarily concerned that ISPs will charge them discriminatory prices for access to reserved bandwidth. When they examine carrier plans for deployment of fiber, they worry that carriers will sell reserved bandwidth to selected application providers rather than adding capacity to their Internet access offerings for shared use by all Internet traffic. The application providers are large business customers of ISPs, and they consider it acceptable for an ISP to charge their *own customers* for access to the Internet based on bandwidth. However, the application providers do not want to be forced to make contractual arrangements with the ISPs of each of the *application provider's customers*. They argue that ISPs have always charged only their *own customers*, not customers of *peering ISPs*. They are ok with ISPs charging their own customers different prices for different levels of bandwidth. Application providers are also concerned that ISPs will choose to strike exclusive deals for access to reserved bandwidth, tilting the playing field among application providers.

Application providers remain unconcerned that ISPs might refuse to offer access to reserved bandwidth to application providers who compete with the ISPs own service offerings. Such behavior is likely for some ISPs who deploy their own video service, particularly if the available bandwidth is not high enough to support a large number of video streams. Application providers, however, typically believe that net neutrality need only apply to Internet access, and do not attempt to apply to it what they view as private network services[17].

Some application providers are concerned that ISPs will not offer access to QoS mechanisms, or at least not without charging application providers. They worry that ISPs will use QoS to support their own real-time applications but refuse access to QoS to competing application providers' traffic. Alternately, they worry that ISPs will attempt to charge application providers, rather than the ISP's own customers, for QoS.

Application providers are concerned that lack of access to reserved bandwidth or QoS will hinder the development of new applications. They argue that small application providers will not be able to compete effectively.

The application providers want a form of net neutrality that prohibits discrimination on the basis of source, destination, or ownership of Internet traffic. Discrimination on the basis of application type or consumer payment, e.g., enhanced performance for all voice and video traffic, would be allowed. This prohibition would only apply to what is thought of as Internet traffic, as opposed to private network services. Application providers generally believe that such a prohibition would not allow ISPs to charge application providers for reserved bandwidth or QoS.

---

[17] However, as mentioned above, we believe it will be increasingly difficult to determine which services are Internet services and which are private services.

The consumer groups in support of net neutrality are generally Openists. They share many of the same concerns as the application providers. In addition, they are concerned that ISPs may block access to certain applications or websites. Some groups are worried that an ISP may block access to websites based on the content of the website, e.g., political beliefs, and interpret net neutrality as an issue of free speech. They are also worried about new charges that may increase the cost of Internet access.

The anti net neutrality lobby consists mostly of ISPs (led by AT&T, Verizon, and Comcast, but including many other wireline and wireless ISPs). They fall squarely into the Deregulationist camp. They argue that there is no current substantial problem, and hence no need for regulation. Furthermore, they argue that if substantial problems arise, then the FCC has the expertise and authority to effectively take action. The ISPs believe that net neutrality would reduce ISPs' incentives to invest in their network infrastructures by reducing the return they may earn on these investments.

The ISPs further believe that they should be allowed to choose whether or not to make available reserved bandwidth or QoS, to charge consumers for these services, and to make individual contractual arrangements with application providers. They argue that refusal to provide access to reserved bandwidth or QoS to application providers would not hinder the application provider's ability to deploy new services. Most ISPs expect to offer access to reserved bandwidth or QoS to application providers for a fee, but want the flexibility to do this through individual contractual arrangements. Such ISPs are typically focused in the short term on arrangements in which the application provider becomes an ISP business customer, rather than providing access through peering ISPs. Some ISPs also claim that net neutrality impinges on the ISP's ability to perform network management tasks, including traffic shaping for p2p traffic.

Finally, many ISPs fear that net neutrality would reestablish many other provisions of common carrier regulation. In particular, they worry that net neutrality's occasional use of the term *nondiscriminatory* will lead to imposition of other common carrier rules such as unbundling and open access.

3.3   Bills

In 2005-2006, Congress considered several approaches to net neutrality. The House bill, the *Communications Opportunity, Promotion, and Enhancement Act of 2006*, and the Senate Commerce Committee bill, the *Advanced Telecommunications and Opportunities Reform Act*, take similar approaches to net neutrality and essentially present the anti net neutrality position. They give the FCC authority to enforce the FCC's four principles (discussed below), but do not allow the FCC to issue additional rulemaking to further define or extend these principles. Blocking of web pages or applications is prohibited, but no further restrictions are placed on ISPs. ISPs remain free to offer or deny access to QoS mechanisms. Either degraded service or enhanced service can be based on application type, source, destination, consumer payment, and/or provider payment. Payments can be on the basis of commercial arrangements, meaning that there is no requirement that they be consistent or equitable between different application providers.

The *Internet Non-Discrimination Act of 2006* (sponsored by Sen. Wyden) essentially presents the most extreme pro net neutrality position among the bills. It prohibits blocking of web pages and applications. The bill includes a broad prohibition on discriminatory behavior; our interpretation is that any use of QoS is prohibited, since the bill prohibits discrimination in data transport, which presumably even includes differentiation based on application type. The bill explicitly prohibits charging application providers that are not customers of the ISP, although this seems redundant given the prohibition on discrimination. The bill prohibits degradation of traffic; our interpretation is that traffic shaping would fall into this prohibition. (These provisions are summarized in table 1.)

	House & Senate bills	Wyden bill	Markey & Sensenbrenner & Snowe-Dorgan bills	Snowe-Dorgan amendment
Prohibit blocking	Y	Y	Y	Y
Degradation:				
allow traffic shaping	Y	N	Y	Y
prohibit if not based on type	N	Y	Y	Y
QoS:				
allow based on type	Y	N	Y	Y
prohibit exclusive use	N	n/a	Y	Y
QoS fees:				
allow consumer payment	Y	n/a	N	Y
allow provider payment	Y	n/a	N	?
prohibit discriminatory fees	N	n/a	n/a	?
prohibit unreasonable fees	N	n/a	n/a	?

Table 1. Comparison of the net neutrality provisions in various Congressional bills

Three other bills – the *Net neutrality Act of 2006* (sponsored by Rep. Markey, defeated as an amendment to the House bill), the *Internet Freedom and Nondiscrimination Act of 2006* (sponsored by Rep. Sensenbrenner), and the *Internet Freedom Preservation Act* (sponsored by Sen. Snowe and Sen. Dorgan) - take similar approaches to each other and present refined pro net neutrality positions. They each prohibit blocking of web pages and applications, as do the other bills. They each also include a broad prohibition on discriminatory behavior; our interpretation is that use of QoS is allowed if based on the type of application, but prohibited if based on the source or destination of the traffic, e.g., based on the application provider. The bills explicitly prohibit charging for QoS, either in the form of customer or application provider payment. The bills also prohibit degradation of traffic on the basis of source or destination, but allow network management based on application type, e.g., traffic shaping.

Finally, Senators Snowe and Dorgan proposed a streamlined version of their bill as an amendment to the Senate bill; this amendment was defeated in the Committee markup on a tie vote. The amendment states a broad prohibition on behavior that discriminates on the basis of source, destination, or ownership of Internet traffic. This has the effect of prohibiting blocking of web pages and applications,

allowing QoS or degradation on the basis of the type of application, and prohibiting QoS or degradation on the basis of the source or destination of the traffic. The amendment is silent on whether an ISP can charge for QoS. Our interpretation is that it allows consumer payment, but it is less clear whether it allows provider payment and if so whether such payment would be mandated to be reasonable and nondiscriminatory.

4.   A Framework for a New Net Neutrality Policy

4.1   Challenges from Convergence

U.S. federal communications policy was separately developed for telephone networks, cable video networks, and cellular networks. As others have noted, convergence between voice, video, and data services is therefore a fundamental challenge to such *vertical regulation*. Common carrier regulation has focused on voice service, and does not effectively address video or data service. Similarly, cable services regulation has focused on video service, and does not effectively address voice or data service. Internet services are not effectively addressed anywhere in the Communications Act. Furthermore, when voice, video and data services are offered over a integrated converged network infrastructure, it is unclear how to apply conflicting regulations from different titles.

This vertical regulation approach started breaking down decades ago, as the telephone network progressed from a set of local monopolies to a somewhat competitive marketplace. Since the Telecommunications Act of 1996, services directly based on the infrastructure are classified as *telecommunication services*, and applications are classified as *information services*. This approach is an attempt at *horizontal regulation*.

The distinction between telecommunication services and information services has been an effective tool for communications policy. However, convergence has now progressed to the point where this distinction is no longer sufficient. Indeed, many of the topics in the 2006 communications bills stem from convergence, including video franchising, net neutrality, VoIP, municipal broadband, universal service, broadcast flag, and white spaces. Despite this trend, Congress has yet to consider creation of a communications policy that fundamentally recognizes convergence.

In the absence of explicit statutes regarding Internet access or services, the FCC recently addressed the issue of whether Internet access is a telecommunications service or an information service. The decision by the FCC to classify Internet access (including DSL and cable modem service) solely as information services exempted these services from regulation under Title II of the Act, and left them only subject to the FCC's Title I ancillary authority. This decision means that Internet access is not subject to the requirement that charges and practices be just and reasonable (section 201), the prohibition on unjust and unreasonable discrimination (section 202), or the requirement that interconnection rates, terms, and conditions be just, reasonable, and nondiscriminatory (section 251). It is this decision that inspired the pro net neutrality coalition.

In 2005, the FCC issued a set of principles related to net neutrality. They express the sentiment that consumers should be entitled to connect devices and to access content and applications of their choice. The principles also state that consumers are entitled to competition among network and application providers. However, the principle regarding competition is too vague to be useful, and the principles are not binding.

4.2   Our Approach

We begin by strongly disagreeing with the FCC's decision to classify Internet access solely as an information service. Telecommunication services can be thought of as the lower layers of the network, and information services can be thought of as the upper layers. The immediate consequence of this approach, purely on a technical basis, is that the Internet clearly consists of both telecommunication services and information services.

We find it useful to separate the layers into two subsets. We consider OSI layers 1-3 (the Internet LAN-link and network layers) as network infrastructure and OSI layers 4-7 (the Internet transport and application layers) as network applications.

A number of other papers have suggested using layers as a tool to formulate communications policy. Lessig (2001) considers a model consisting of physical, logical, applications, and content layers. He argues that the physical and logical layers of the Internet have historically been neutral. He believes it is acceptable for the physical layer to be closed, but proposes that the logical layer should be open and act as a commons. Werbach (2002) uses a similar set of four layers: the physical layer corresponds to OSI layer 1, the logical layer to OSI layers 2-6, and the application and content layers share OSI layer 7. He argues that communications policy should be formulated around these layers, with open interfaces between them. Solum and Chung (2003) propose a six-layer model, and argue that communications policy should attempt to respect the integrity of layers and to place regulation at or near the layer where the problem occurs. Whitt (2004) suggests a four-layer model similar to Werbach's, and presents principles concerning how layers should inform policy formulation.

While all of these approaches support horizontal regulation, and many result in a wider applicability than considered here, most of these layered models combine OSI layers 3 and 4 (the Internet network and transport layers) into a single logical layer. We believe that these two OSI layers display quite different characteristics. The infrastructure layers (OSI layers 1-3) include functions that must be implemented *in every portion of the network* including the access network, e.g., routing, addressing, and QoS (if used). The application layers (OSI layers 4-7), in contrast, include functions that can be implemented *at any point within the network*, e.g., e-mail, web-hosting, and caching.

One of the strengths of this delineation is that the network infrastructure layers exhibit a *high barrier-to-entry*, while the application layers exhibit a *low barrier-to-entry*. The high barrier-to-entry of infrastructure has heretofore resulted in a small number of carriers offering service in any particular

location. In contrast, the low barrier-to-entry of applications has resulted in a competitive market with a large number of application providers.

The distinction between network infrastructure and applications also helps with correct application of the end-to-end principle. Using the new terminology, we can view the principle as a suggestion that network functionality should be implemented in network infrastructure *only if* it cannot be implemented effectively in network applications. With the recognition that OSI layers 1-3 are implemented *in every hop* along the transmission path, whereas OSI layers 4-7 are only implemented *in the end devices*, this is equivalent to the suggestion that network functionality should be implemented in every hop only if it can not be implemented effectively solely at the endpoints.

With the Internet's layered architecture in mind, we now turn to the Communications Act for potential policy models. We then meld these models with our understanding of the architecture and functionality of the Internet to form broad policy goals.

The Communications Act broadly lays out three potential models to follow. First, if a communications market has sufficient competition, then minimal regulation is applied. This is captured in Title I of the Act, which applies forbearance from regulation unless required to ensure just and reasonable practices. The paradigm here is that a free market will regulate itself, and will promote investment in a fair manner that maximizes social welfare.

If sufficient competition is not present to ensure just and reasonable practices, Titles II and VI apply different models of regulation. Both models were initially formulated assuming a local monopoly over a portion of the corresponding network, but have since been applied to circumstances in which there is limited (but not sufficient) competition. Both models attempt to maximize social welfare by limiting monopoly behavior.

Title II (common carrier) uses the open access model. The network can be separated into two layers: infrastructure and applications. Local infrastructure, e.g., the local loop, is usually deemed to be lacking sufficient competition, due to the high cost of building this infrastructure. Applications, e.g., telephone service, in contrast are often offered by a large number of competitors. Title II essentially dictates open access to infrastructure, and requires that carriers make critical network elements available to application providers without unjust or unreasonable discrimination and without undue or unreasonable preference. Title II also mandates reasonable and nondiscriminatory rates, terms, and conditions for interconnection with other infrastructure providers. As a result, telephone carriers do not limit who a customer can call and do not provide differentiated service based on who a customer calls.

Title VI (cable services), in contrast, uses a paradigm of a closed network. In this model, the network cannot be naturally separated so that infrastructure and applications can be offered by separate entities. Vertically integrated carriers control both the infrastructure and the applications. Title VI recognizes this natural vertical integration, and applies limits to the cable carrier's behavior when it is also

integrated with content providers. As a result, cable carriers do determine which content a customer can receive.

We believe that the Internet's architecture supports the need for some type of openness. However, we do not believe that net neutrality by itself requires open access. This intermediate approach places us loosely in the Nondiscriminationist camp. We suggest below that net neutrality can be effectively achieved by a properly regulated *open interface* from applications to network infrastructure. The requirement of an open interface captures the central tenet of a layered Internet architecture, and yet is less intrusive than the requirement of open access.

We also believe economic arguments support the need for some type of openness. We believe that carriers should offer an interface to network infrastructure to applications without unjust or unreasonable discrimination and without undue or unreasonable preference. We argue below that an open interface mandate can be used to limit monopoly or oligopoly behavior. We adopt the free market model, based on Title I forbearance, for the application portion of the Internet. We mandate an open interface, based on selected elements of Title II, for the local infrastructure portion of the Internet. We believe this combination is most likely to encourage continued development of the Internet in a manner that maximizes social welfare.

We believe that an open interface should attempt to provide a level playing field between application providers and ISPs who offer competing applications. Threats to a level playing field are potentially posed by vertical integration between Internet service providers and application providers. Certain *Internet applications* (e.g., video on demand, high quality VoIP, video conferencing) require specialized *Internet infrastructure services* (e.g., dedicated bandwidth or QoS). Our goals for net neutrality are:

> **Goal A)** ISPs should be prohibited from refusing to provide enabling Internet infrastructure services to competing application providers in order to differentiate the ISP's own application offerings,
>
> **Goal B)** ISPs should be prohibited from providing Internet infrastructure services to competing application providers at inflated prices in order to favor the ISP's own application offerings, and
>
> **Goal C)** ISPs should be prohibited from making exclusive deals to provide enabling Internet infrastructure services to certain application providers.

Such threats to a level playing field are limited in scope, and therefore any remedy should be similarly limited in scope. First, vertical integration only affects applications that rely on specialized Internet infrastructure. A characteristic of such network services is that they must be provided directly in the *access network* (e.g., "the last mile"). Networks services that can be provided outside the access network (e.g., caching, search engines, web hosting) can be provided by competing ISPs or application

providers, and therefore do not represent a threat to a level playing field. Second, when sufficient competition exists in the Internet access market, there is a lower danger that vertical integration would cause a non-level playing field.

Our approach, therefore, is to apply existing elements of the Communications Act that prohibit anti-competitive and unreasonably discriminatory behavior to ensure a level playing field. Our goal is to apply these elements, however, in a limited fashion:

Goal D) Limit the scope through new definitions that transcend the limitations of current telecommunication and information services definitions,

Goal E) assure forbearance where sufficient competition exists,

Goal F) ensure that ISPs have the right to apply network management mechanisms that do not threaten a level playing field, and

Goal G) ensure that regulation does not impede an ISP from making arrangements with consumers, application providers, and peering ISPs for Internet infrastructure services in a manner that does not conflict with the above goals.

4.2   Comparison to other approaches

Openists believe that the Internet is best served by maintaining a dumb network that does not differentiate among different types of traffic. They use the end-to-end principle to support this position. We disagree with both the reasoning and the conclusion. We believe that the end-to-end principle suggests implementing functionality in the application layers if feasible. However, as noted above, QoS mechanisms such as priority or reservations must be implemented at the Internet LAN-link and/or network layers in order to be effective. The end-to-end principle therefore does not ban their use, and therefore does not mandate a dumb network. Indeed, one of the original authors of the end-to-end principle was also one of the authors of an Internet protocol to implement QoS using reservations (Braden, Clark, and Shenker, 1994), and has since suggested that a new theory addressing network core functionality should be created to co-exist with the end-to-end principle (Blumenthal & Clark, 2001) and that the end-to-end principle should be generalized with a new form of modularity that can gracefully accommodate tussle along various competing players (Clark et. al., 2005). We believe that the introduction of QoS, and hence the creation of a smart network, will *increase* the range of applications that can be efficiently run on the Internet.

Openists support a policy based on open access, including a broad ban on vertical integration. We agree with the need for an open interface between network infrastructure and applications, but do not believe that net neutrality by itself is threatened by vertical integration that crosses other layer boundaries.

Deregulationists present economic arguments that regulation will hinder investment by ISPs and that in the absence of regulation ISPs will vertically integrate only in ways that maximize consumer welfare. However, other economists argue that the same theory of network externalities indicates that ISPs do have an incentive to discriminate against competing application providers, and that the economic benefits of net neutrality outweigh the costs. We accept here this latter claim, and agree with the goal of providing a level playing field. Although we agree that development of applications drives network infrastructure investment, we do not believe that vertical integration violating an open interface between infrastructure and applications is necessary to foster new applications.

Some deregulationists have also suggested that competition can be more effectively fostered between different physical networks that are matched to different types of applications. We strongly disagree, and believe that 20 years of convergence provides strong evidence for the efficiency of integrated networks over the use of separate infrastructures.

Finally, many deregulationists equate QoS with other types of discriminatory network behavior (e.g., caching or priority listing of search results) and use this to argue that QoS should not be regulated. We strongly disagree. We are concerned with the use of prioritization when there is limited competition (e.g., QoS); we have faith that free market dynamics will adequately address the use of prioritization when there is sufficient competition (e.g., caching or priority listing of search results). Equating the two types of prioritization ignores the distinction between network infrastructure services (e.g., QoS) which must be implemented in the access network and network applications (e.g., caching) which can be implemented in other ISPs' networks.

We agree with the Nondiscriminationist camp's arguments that there are good and bad uses of traffic discrimination. Although we do not believe that QoS requires vertical integration, we support a policy that allows vertical integration and traffic differentiation, but restricts their use to ensure that ISPs do not discriminate in a manner that extracts oligopoly rents. However, whereas Wu (2003) would ban discrimination at the Internet network layer and hence ban QoS based on the type of application, we believe that an open interface is a cleaner and more effective solution. This difference will become clear as we analyze what types of ISP network management should be considered unacceptable.

5.   Acceptable and Unacceptable ISP Network Management

In this section, we use our approach to delineate which types of network management should be considered acceptable and which unacceptable. We consider a wide range of network management actions related to net neutrality. For each, we decide whether the action should be allowed or prohibited by considering the goals presented in the last section as well as the general intent of the Communications Act as applied to other technologies. We focus on the case in which sufficient competition does not exist in the Internet access market; where sufficient competition exists we choose to apply forbearance. When doubt remains as to the desirable policy goal, we choose the action most likely to encourage a level playing field.

5.1   Blocking

First, consider *blocking* of either a web page or an application. Currently, blocking (when used) is usually done on the basis of the *type of application*. This type of blocking can be beneficial, e.g., firewalls that block security intrusions. However, it can also be anti-competitive, e.g., an ISP that blocks all VoIP traffic in order to limit competition to its own voice service. In some instances, blocking based on application type can be viewed as either beneficial or harmful depending on user preferences, e.g., a firewall that blocks all file-sharing traffic.

Is it acceptable for an ISP to block applications, on the basis of the application type? In the absence of sufficient competition in infrastructure, we believe such blocking is not acceptable because it may violate goal A stated above. While some ISPs block certain applications to limit bandwidth use, we believe that network management should be transparent and that tiering is a superior solution. Even if sufficient competition exists, then the layered architecture of the Internet strongly suggests that such blocking should be rare. Our answer is that blocking based on application type should be a decision made by the consumer, not the ISP.

It is less common to block web pages or applications based on the *source and/or destination*. However sometimes this type of blocking is used for security purposes. Examples include parental control software and firewalls that block all traffic from unknown addresses. Is it acceptable for an ISP to block applications, on the basis of the source and/or destination of the traffic? We believe this is fairly straightforward, and that blocking based on source and/or destination should also be a decision made by the consumer, not the ISP.

5.2   Degradation

Next, we consider *degradation* of Internet traffic. Currently, intentional degradation is common in a few scenarios. First, it is common to place a limit on the bandwidth of a broadband user's access link that is lower than the speed the link could accomplish. Such limits, referred to as *tiering*, are often used to differentiate service offerings by a selection of access rates, and to charge different prices for these different service offerings. Although tiering is a form of intentional degradation, neither the pro or anti net neutrality lobbies oppose its use. Second, it is increasingly common for an ISP to place bandwidth limits on certain classes of applications, e.g., traffic shaping. Currently, the most common such limits are on file sharing traffic, which otherwise may consume a large proportion of the available bandwidth and cause congestion for all other applications. The bandwidth limits degrade the performance of the applications to which they are applied, but do not block the applications entirely. Both tiering and traffic shaping are intentional degradation based on the type of application (and in the case of tiering on consumer payment). In contrast, degradation of traffic based on the source and/or destination is both technically difficult and uncommon[18].

---

[18] We do not consider congestion control measures (e.g., TCP) which treat all users similarly as being intentional degradation based on source or destination.

Is it acceptable for an ISP to intentionally degrade traffic? We believe that degradation based on application type, e.g., tiering and traffic shaping, is acceptable since it does not bias the network toward certain providers. Indeed, we believe that such techniques can be crucial elements of network management, support goal F, and should not be discouraged. However, we believe that any degradation based on source or destination should also be a decision made by the consumer, not the ISP, since it would violate goal A in much as the same manner as blocking based on source or destination.

5.3   QoS

Finally, we consider QoS mechanisms such as reservations or priorities. QoS is often used by carriers to provide acceptable performance for their own VoIP or video conferencing traffic as part of business packages that offer combined voice, video, and data transport. Although use of QoS for residential service is currently less common, we expect its use to increase with the growth in residential VoIP, video conferencing, and video streaming.

A traffic stream receives enhanced performance through QoS if the stream requests it and the carrier grants the request. A carrier may grant QoS requests: (1) to only its *own applications* (e.g., only its own VoIP subscribers), (2) to *particular applications types* (e.g., VoIP packets from all providers), or (3) on the basis of *contractual arrangements* (e.g., consumer and/or application provider payment).

A central component of net neutrality policy is the decision of how to treat QoS mechanisms. Is it acceptable for an ISP to implement QoS in its network? Some net neutrality proponents argue that the Internet should only provide best-effort service so that all Internet streams are treated identically; this group opposes any use of QoS. Although some carriers may choose alternate methods to provide acceptable performance to voice and video applications, we believe that a broad ban on use of QoS is a Luddite instinct. Furthermore, we do not see any precedent in the Communications Act for a similar ban on a set of technologies. We therefore believe use of QoS can be acceptable. The real issue, we believe, is how QoS is used.

Is it acceptable for QoS to be based on application type? Currently, the Internet adopts different types of transport[19] for different types of applications. QoS would fall into the same category. We do not see a threat to goals A-C through this use of QoS, and our open interface approach would allow it. As an example, if a carrier chooses to deploy QoS mechanisms in their network and to make these mechanisms available to all VoIP traffic, regardless of the application provider supporting the service, we believe this approach should be allowed. The anti net neutrality lobby supports this position. The pro net neutrality lobby, excepting those who believe QoS should not be allowed in any manner, also supports this position.

---

[19] TCP versus UDP.

Is it acceptable if a carrier deploys QoS in its network to support some of its own applications, but does not offer these mechanisms to competing application providers? The anti net neutrality lobby argues that refusal to provide access for QoS to application providers would not hinder the application provider's ability to deploy new services. We disagree. QoS should be classified as infrastructure, not application, because it must be implemented *in every portion of the network* in which congestion occurs in order to be effective. As a result, there is no effective option available to application providers who wish to provide enhanced performance to voice and video applications; they must receive the same QoS as the carrier uses for its own similar applications in order to compete effectively. In support of goals A and C, we conclude that it is not acceptable for a carrier to deploy QoS in its network to support some of its own applications, but not to offer these mechanisms to competing application providers. Therefore, we believe that QoS on the basis of source or destination is not acceptable.

The remaining issue with use of QoS is whether it is acceptable for QoS to be based on consumer and/or application provider payment. The anti net neutrality lobby argues that this approach should be allowed without constraint. Such contractual arrangements based on consumer payment are likely to be combined with tiering, e.g., for a small additional fee all VoIP traffic to/from a particular residence will receive QoS. Alternately, contractual arrangements for QoS could be based on provider payment, e.g., all VoIP traffic to/from a particular provider will receive QoS. Both approaches are technically feasible. The fee for such service could be a flat monthly fee, a per-minute charge, or a per-byte charge. Any consumer fee would almost certainly be based on a publicly available fee schedule, and therefore be nondiscriminatory. We see no threat to goals A through C from QoS based on consumer payment, and believe allowing this approach supports goal G. We also note that most pro net neutrality groups would agree.

Application provider fees for QoS, however, could potentially be subject to *commercially negotiated* agreements, as are commonly used when carriers offer virtual private networks to business customers. Such agreements open the door to potential discrimination between various application providers. Under the policy developed above, we have already chosen to prohibit an ISP from offering QoS to only selected application providers, so the issue here is whether an ISP can charge different fees to different application providers. We believe the open interface approach suggests that any such discrimination must be reasonable, e.g., based on the differences in cost of providing QoS. We conclude that similar constraints should be placed on QoS based on provider payment.

5.4   Summary

The decisions made in above with regard to what types of network management should be allowed or prohibited result in an answer of yes for every row in table 2. All of the bills in Congress prohibit blocking, as would we. With regard to degradation, our decision to allow traffic-shaping puts us at odds with the Wyden bill, whereas our decision to prohibit degradation on the basis on source or destination (unless elected by the consumer) puts us at odds with the House and Senate bills. On use of QoS, our decision to allow QoS based on application type puts us at odds with the Wyden bill, whereas our decision to prohibit exclusive use of QoS puts us at odds with the House and Senate bills. Finally, our decision to allow either consumer or provider payment for QoS puts us at odds with the Markey,

Sensenbrenner, and Snowe-Dorgan bills, whereas our mandate that such fees be reasonable and nondiscriminatory puts us at odds with the House and Senate bills.

To illustrate these policy decisions, consider their effect upon an ISP who offers both broadband Internet service and a separate VoIP service. The effects of our policy goals are shown in table 2.

Is it acceptable if …	Answer
… a carrier *blocks* a competitor's VoIP traffic?	no
… a carrier doesn't block a competitor's VoIP traffic, but doesn't offer QoS to *competitor's* VoIP subscribers while using QoS for its *own* VoIP subscribers?	no
… a carrier gives a *broadband subscriber* who uses a competitor's VoIP service the choice of (a) best-effort transport of their VoIP traffic as part of the basic broadband package, (b) enhanced performance for their VoIP traffic for an additional 1¢/min paid by the subscriber, or (c) enhanced performance for up to 500 minutes of their VoIP traffic for an additional $5/month paid by the subscriber?	yes
… a carrier gives a *VoIP competitor* the choice of (a) best-effort transport of their VoIP traffic as part of the subscriber's basic broadband package or (b) enhanced performance for their VoIP traffic for an additional 1¢/min paid by the VoIP provider?	yes
… a carrier charges different VoIP competitors *different prices* for QoS?	no
… a carrier charges VoIP competitors a uniform price for QoS, but a different price than charged to *its own affiliates*?	no

Table 2. Effect of policy goals upon an ISP offering VoIP.

6.   Statute Language

Goals A-G, as presented above, are implemented in draft statute language in the following subsections. The new definitions of Internet infrastructure services, Internet application services, and access network are presented in section 6.1, which helps implement goal D. Sections 6.2-6.4 guarantee consumers the right to an open interface between infrastructure and applications, implementing goals A-C. Section 6.5 guarantees ISPs the right to use network management techniques in a pro-competitive manner, implementing goals E-G.

In each section, we alternate between presenting draft statute language (indented and written in *italics*) and discussion of the statute language.

6.1   Definitions

As discussed above, we strongly disagree with the FCC decision to classify Internet access solely as an information service. We present new definitions of Internet infrastructure and applications to rectify

this, and to introduce terms that are meaningful in both telephone networks and the Internet. These terms will be used as a foundation for the language in the following sections:

> *SECTION 1. DEFINITIONS.*
> *In this title:*
>
> *(1) INTERNET INFRASTRUCTURE SERVICES. The term 'Internet infrastructure services' means all services- (A) over a network that uses a public right-of-way; and (B) that reside at or below the network layer or are required to manage the network.*
>
> *(2) INTERNET APPLICATION SERVICES. The term 'Internet application services' means all services-(A) over a network that uses a public right-of-way; (B) that are not infrastructure services; and (C) that do not fall under Title VI of the Communications Act.*
>
> *(3) NETWORK LAYER. The term 'network layer' means the third layer of the 7-layer Open Systems Interconnection Model, responsible for message addressing and for routing information within the network, including routing within the telephone network and including the Internet Protocol within the Internet.*

These terms agree with the intent of the older terms *telecommunication services* and *information services*. Internet infrastructure services do not change the content of information, similar to telecommunication services. In contrast, Internet application services create, store, or change the presentation of information, similar to *information services*.

Maintaining a distinction between the Internet infrastructure and applications is critical to formulation of good communications policy. Internet infrastructure services can only be provided by carriers, and must be provided by each carrier on their portion of the network. Such services include multi-user sharing of a wire or frequency (e.g., Ethernet or WiFi), routing (e.g., IP), and address assignment (e.g., DHCP). Internet infrastructure services require large investments into loops or wireless spectrum and switches or routers. Since Internet infrastructure services often require substantial hardware and must be implemented in each portion of the Internet, they reflect a high barrier-to-entry. Internet application services can be provided by carriers or by many other application providers on the Internet. Such services include e-mail, web-hosting, caching, voicemail, and the portions of VoIP and IPtv that can be offered by independent application providers. Since Internet application services are typically software-based and can be placed at many locations within the Internet, they usually reflect a low barrier-to-entry.

Among Internet infrastructure services, a further geographic distinction is worthwhile. *Access networks*, commonly called *the last mile*, are the portions of the Internet that must be transversed to communicate with other Internet locations. Access networks currently often have limited competition. The portion of the Internet that resides outside of access networks is commonly called the *Internet backbone*.

The backbone currently exhibits substantial competition. Section 1 presents a definition of access networks:

> **(4)** ACCESS NETWORK -The term 'access network' means the portions of the Internet service provider's network which must be transversed to form routes from the Internet to its subscribers.

These definitions effectively limit the scope of regulation presented in other sections, and implement goal D.

6.2   Principles

Many of the bills addressing net neutrality incorporate some of the FCC principles. We start by implementing the first three FCC principles, with language based roughly on the Senate Commerce bill. These principles guarantee a subscriber's right to use and offer content and applications, and guarantee a subscriber's right to connect devices.

> SECTION 2. PRINCIPLES
>
> Except as otherwise provided in this title, with respect to Internet services, each Internet service provider shall allow each subscriber to-
>
> **(1)** access, use, post, receive, or offer any lawful content, application, or service of that subscriber's choosing;
>
> **(2)** connect any legal device of that subscriber's choosing to the Internet access equipment of that subscriber, if such device does not harm the network of the Internet service provider;

We next implement part of the fourth FCC principle, with language taken from the Snowe-Dorgan amendment to the Senate Commerce bill. Unlike the Snowe-Dorgan amendment, however, this provision is subject to allowed network management techniques as specified in section 5. In conjunction with section 5, this guarantees ISPs the right to implement bandwidth allocation and QoS and to sell such service, and it guarantees subscribers the right to purchase such service that discriminates only on application type, broadly supporting goals A-C.

> **(3)** receive service that does not discriminate in the carriage and treatment of Internet traffic based on the source, destination, or ownership of such traffic; and

Finally, we implement a often proposed 5th principle, the consumer's right to information, using language taken from the Senate Commerce bill:

(4) *receive clear and conspicuous information, in plain language, about the estimated speeds, capabilities, limitations, and pricing of any Internet service offered to the public.*

6.3   Discrimination

Although principle *(3)* above is an elegant statement of the pro net neutrality lobby goals, we do not believe it is sufficient to fully implement goals A-C. In particular, we do not believe that it alone prohibits unjust or unreasonable discrimination in charges. To reinforce these goals, section 3 of the draft statute language applies to ISPs, in a restricted manner, the portion of the Title II of the Communications Act that places limits on use of discrimination and preferences:

SECTION 3. DISCRIMINATION AND PREFERENCES

For purposes of sections 202 and 206 through 209 of the Act, an Internet service provider shall be treated as a common carrier, and Internet infrastructure service shall be treated as a communications service.

We apply the limits specified in section 3 only to services that can be provided only by the ISP, using the new term *Internet infrastructure service*. We also apply these limits only when a competitive market is absent, through coordinating provisions in section 4 below. Section 3 does not apply other requirements on common carriers, e.g., tariffs, interconnection, or unbundling, to ISPs.

6.4   Competition

The Internet has developed in a manner in which ISPs often also serve as application providers. To directly address such vertical integration, we draw upon other provisions from Titles II and VI of the Communications Act that address vertical integration in telephone and cable networks.

First, however, we implement goal E by limiting application of net neutrality to situations in which sufficient competition does not exist. This can be accomplished by ensuring that the forbearance provisions currently in Title I of the Communications Act apply to ISPs offering Internet infrastructure service:

SECTION 4. COMPETITION

(a) *For purposes of section 10 of the Act, an Internet service provider shall be treated as a telecommunications carrier, and Internet infrastructure service shall be treated as a telecommunications service.*

This provision therefore allows forbearance from the limits placed through section 3 when there is a competitive market.

We turn next to the responsibilities of vertically integrated ISPs to *their own subscribers and peers*. When an ISP offers applications that rely on Internet infrastructure services, we require the ISP to make available to competitors the same Internet infrastructure services at the same prices:

> **(b)** *An Internet service provider shall make available to subscribers and other Internet service providers - on the same prices, terms, conditions of sale, and delivery - any Internet infrastructure services provided on its access networks as the Internet service provider offers to Internet application services provided by itself or its affiliates.*

> **(c)** *An Internet service provider shall provide Internet infrastructure service to subscribers and other Internet service providers, that is at least equal in quality to that provided by the Internet service provider to itself or its affiliates.*

Provision (b) is based on section 628(c)(2) of the Communications Act, which governs program access in vertically integrated video programming and delivery. Provision (c) is based on section 251(c)(2)(c), which addresses quality of service responsibilities for interconnecting carriers.

Finally, we turn to the responsibilities of vertically integrated ISPs to *subscribers of other ISPs*. Without regulating peering agreements, we wish to prohibit a vertically integrated ISP from using service level agreements with its peers to favor its own applications. We use language based on section 628(b) of the Communications Act, which places similar restrictions on vertically integrated video programming and delivery:

> **(d)** *It is unlawful for an Internet service provider to engage in unfair methods of competition, unreasonably discriminatory conduct, or unfair or deceptive acts or practices, the purpose or effect of which is to hinder significantly or to prevent any Internet application provider from providing content, applications, or services to consumers.*

Provisions (b), (c), and (d) reinforce goals A and B.

6.5   Network management

Goal F requires a delineation of allowed network management practices. The Senate Commerce bill included most of the required provisions:

> *SEC. 5. NETWORK MANAGEMENT.*

An Internet service provider may-

(1) *protect the security, privacy, or integrity of the network or facilities of such provider, the computer of any subscriber, or any service, including by- (A) blocking worms or viruses; or (B) preventing denial of service attacks;*

(2) *facilitate diagnostics, technical support, maintenance, network management, or repair of the network or service of such provider;*

(3) *prevent or detect unauthorized, fraudulent, or otherwise unlawful uses of the network or service of such provider;*

(4) *block access to content, applications, or services that Federal or State law expressly authorizes to be blocked, including child pornography;*

(5) *provide consumers Parental Control applications, devices, or services, including- (A) blocking access to websites with obscene or adult content; (B) blocking display of video content based on a common rating; or (C) offering a family friendly tier of service;*

(6) *allow a subscriber to elect to have content, applications, or services blocked at the request of such subscriber;*

We add a 7th provision to explicitly guarantee ISPs the right to alleviate congestion by treating all traffic similarly, or by treating all applications of the same type similarly:

(7) *alleviate congestion in a manner that does not distinguish based on the source or ownership of content, application, or service;*

Finally, we suggest two provisions to support goal G. Both work in conjunction with provision *(3)* in section 2. Provision *(8)* guarantees ISPs the right to sell reserved bandwidth and QoS to both their residential and business subscribers. Provision *(9)* guarantees ISPs the right to discriminate in the carriage of Internet traffic based on peering arrangements with other ISPs. Such peering arrangements would allow for reserved bandwidth and QoS to be provided cooperatively by multiple ISPs.

(8) *offer directly to a subscriber Internet service at different prices based on defined levels of bandwidth, quality of service, or the actual quantity of data flow over a user's connection, and discriminate in the carriage and treatment of Internet traffic based on such contract with that subscriber; and*

(9) *enter into contracts with other Internet service providers, and discriminate in the carriage and treatment of Internet traffic based on such contract with that Internet service provider.*

6.6   Further Study

Both the House and Senate bills required the FCC to conduct further review related to net neutrality. We suggest here a study more tightly focused on potential anti-competitive problems, using the new definition of Internet infrastructure services:

SECTION 6. FCC REVIEW

(a) *IN GENERAL. Beginning 1 year after the date of enactment of this Act, the Federal communications commission shall report annually to the Committee on Commerce, Science, and Transportation of the Senate and the Committee on Energy and Commerce of the House of Representatives regarding-*

> **(1)** *the developments in Internet infrastructure services, including Internet architecture, peering and interconnection, routing, and quality of service;*

> **(2)** *contracts between Internet service providers and subscribers for Internet service, including levels of bandwidth and quality of service, and any discrimination in the carriage and treatment of Internet traffic based on such contracts;*

> **(3)** *contracts between interconnecting Internet service providers, and any discrimination in the carriage and treatment of Internet traffic based on such contracts; and*

> **(4)** *how such developments impact the development of Internet applications that rely on evolving Internet infrastructure services.*

(b) *DETERMINATIONS AND RECOMMENDATIONS. The Federal Communications Commission shall make such recommendations under subsection (a), as the Commission determines appropriate.*

7.   Conclusion

We have developed a net neutrality policy based on a delineation of Internet infrastructure and Internet applications. We argued that the net neutrality issue was caused by technology convergence, and that any solution should mirror the architecture of the Internet. We believe that the central problem is the ability of an ISP with market power to use Internet infrastructure to favor its own applications or otherwise discriminate in a manner that extracts oligopoly rents. We illustrated how communications policy can prohibit such anti-competitive behavior without restricting desirable forms of network management. This approach differs with many of the pro net neutrality bills in that it allows nondiscriminatory network management techniques and QoS and it allows consumer or provider payment for QoS; it differs with many of the anti net neutrality bills in that it prohibits discriminatory use of Internet infrastructure and discriminatory degradation and mandates reasonable and nondiscriminatory fees for QoS.

Our proposed statute language is not complete. In particular, enforcement must be addressed. Furthermore, net neutrality is not the only issue resulting from a fragmented communications policy; a broader consideration of the effect of technology convergence upon communications policy is warranted.

References

Bar, F., Cohen, S. S., Cowhey, P., DeLong, B. J., Kleeman, M., & Zysman, J. (2000). Access and innovation policy for the third-generation Internet. *Telecommunications Policy*, *24*, 489-518. (Available at http://repositories.cdlib.org/brie/BRIEWP137).

Bar, F., & Sandvig, C. (2000). Rules from truth: Communication policy after convergence. In *Proceedings of the 28th Telecommunications Policy Research Conference.*

Bloomberg News (2006). AT&T, Comcast Rout Google in `Net Neutrality' Battle. *Bloomberg News, July 20, 2006.*

Blumenthal, M. S., & Clark, D. D. (2001). Rethinking the design of the internet: the end-to-end arguments vs. the brave new world. *ACM Transactions on Internet Technology*, *1* (1), 70-109.

Braden, R., Clark, D., & Shenker, S. (1994). RFC 1633 - Integrated Services in the Internet architecture: An overview. *IETF Network Working Group*.

Clark, D., Wroclawski, J., Sollins, K., & Braden, R. (2005). Tussle in cyberspace: Defining tomorrow's Internet. *IEEE/ACM Transactions on Networking*, *13*, 462-475.

Cox Communications (2004). Cox Communications VoIP Whitepaper. May 2004. (Available at
        http://www.fcc.gov/oet/tac/7.28.04_TAC_Cox_VoIP_whitepaper.pdf).

Farrell, J., & Weiser, P. (2003). Modularity, vertical integration, and open access policies: Towards a
        convergence of antitrust and regulation in the Internet age. *Harvard Journal of Law and
        Technology*, *17*, 85-134. (Available at http://ssrn.com/abstract=452220).

Herman, B. D. (2007). Opening bottlenecks: On behalf of mandated network neutrality. *Federal
        Communications Law Journal*, *59*. (Available at http://ssrn.com/abstract=902071).

Jordan, S. (2007). Implications of Internet architecture upon net neutrality. *ACM Transactions on Internet
        Technology*, *forthcoming*.

Lehr, W. H., Gillett, S. E., Sirbu, M. A., & Peha, J. M.(2006). Scenarios for the network neutrality arms
        race. In *Proceedings of the 34th Telecommunications Policy Research Conference.*

Lemley, M. A., & Lessig, L. (2001). The end of end-to-end: Preserving the architecture of the Internet in
        the broadband era. *UCLA Law Review*, *48*, 925-972. (Available at
        http://ssrn.com/abstract=259491).

Lessig, L.(2001). The Internet under siege. *Foreign Policy*, *127*, 56-65.

McTaggart, C.(2006). Was the Internet ever neutral? In *Proceedings of the 34th Telecommunications
        Policy Research Conference.*

Owen, B. M., & Rosston, G. L. (2003). Local broadband access: Primum non nocere or primum processi?:
        A property rights approach. *Stanford Law and Economics Olin Working Paper No. 263*. (Available
        at http://ssrn.com/abstract=431620).

Peha, J. M. (2006). The benefits and risks of mandating network neutrality, and the quest for a balanced
        policy. In *Proceedings of the 34th Telecommunications Policy Research Conference.*

Saltzer, J. H., Reed, D. P., & Clark, D. D. (1984). End-to-end arguments in system design. *ACM
        Transactions on Computer Systems*, *2* (4), 277-288.

Solum, L. B., & Chung, M. (2003). The layers principle: Internet architecture and the law. *U San Diego
        Public Law Research Paper No. 55*. (Available at http://ssrn.com/abstract=416263).

Speta, J. B. (2000). Handicapping the race for the last mile?: A critique of open access rules for
        broadband platforms. *Yale Journal on Regulation*, *17*, 39-91.

van Schewick, B. (2007). Towards an economic framework for network neutrality regulation. *Journal on
        Telecommunications and High Technology Law*, *5*. (Available at
        http://ssrn.com/abstract=812991).

Werbach, K. D. (2002). A layered model for Internet policy. *Journal on Telecommunications and High-Tech Law*, *1*. (Available at http://ssrn.com/abstract=648581).

Whitt, R. S. (2004). A horizontal leap forward: Formulating a new communications public policy framework based on the network layers model. *Federal Communication Law Journal*, *56*, 587-672.

Wu, T. (2003). Network neutrality, broadband discrimination. *Journal of Telecommunications and High Technology Law*, *2*, 141-178. (Available at http://ssrn.com/abstract=388863).

Wu, T. (2004). The broadband debate: A user's guide. *Journal of Telecommunications and High Technology Law*, *3*, 69-95. (Available at http://ssrn.com/abstract=557330).

Yoo, C. S. (2005). Beyond network neutrality. *Harvard Journal of Law and Technology*, *19*, 1-77. (Available at http://ssrn.com/abstract=742404).

Yoo, C. S. (2006). Network neutrality and the economics of congestion. *Georgetown Law Journal*, *94*. (Available at http://ssrn.com/abstract=825669).

Do wireless networks merit different net neutrality than wired networks?

Scott Jordan

Department of Computer Science, University of California, Irvine

Abstract

This paper attempts to answer whether wireless net neutrality should be different than wired net neutrality because of the different technologies used. We are concerned with whether the challenges of wireless signals and mobility merit different traffic management techniques, and how these techniques may affect net neutrality. We address whether wireless network operators should have a different ability to *restrict applications* used on their networks than wired ISPs. We are also concerned with which wireless applications or services should be covered by a net neutrality requirement, and whether this requires the definition of *managed services*. We review the pertinent aspects of network architecture and discuss the main differences between wired and wireless networks. We find that wireless networks differ substantially from wired networks only at and below the network layer. We then more specifically address how traffic management differs in wireless networks. We conclude that wireless networks require stronger traffic management. We find, however, that despite the differences in traffic management, similar net neutrality concerns apply. We state our general position on net neutrality, and then examine how it may apply to wireless net neutrality. We argue that since the differences between wired and wireless networks lie in lower layers, net neutrality in both wired and wireless networks can be effectively accomplished by requiring an open interface between network and transport layers.

1   Introduction

Net neutrality has typically been discussed in the context of wired networks, with a focus on the wired public Internet. Those who support wired net neutrality generally believe that there is a danger that Internet Service Providers (ISPs) who offer applications may discriminate in favor of themselves over

---

Email address: sjordan@uci.edu (Scott Jordan).

competing application providers. Examples of such conflicts may include cable ISPs that discriminate in favor of their own Voice-over-IP (VoIP) packets over competing VoIP providers' packets and telephone ISPs that discriminate in favor of their own video over IP packets over competing video providers' packets. Those who oppose wired net neutrality generally believe that any such danger does not represent a market failure and that net neutrality regulation will be counterproductive. Good overviews of the arguments on both sides can be found in Clark (2007), Peha (2007), and Jordan (2007).

Recently, the question has arisen over whether and how net neutrality should apply to wireless networks. In wireless networks, similar anticompetitive concerns may apply. Cellular network ISPs may have the incentive to discriminate in favor of their own video packets over competing video providers' packets. In addition, they may have the incentive to discriminate against any applications that compete with their primary revenue streams, including all competing voice and text-messaging applications that run over the Internet Protocol (IP).

The key question that this paper attempts to address is whether wireless net neutrality should be different than wired net neutrality because of the different technologies used. There is little academic literature that directly addresses net neutrality in wireless networks. Wu (2007) started much of the current debate, where he focussed on whether subscribers should be able to attach wireless devices of their choice. Wu argued for the extension of the Federal Communication Commission's (FCC) Carterfone rules FCC (1968) to wireless networks, including a prohibition on locking of devices to a carrier and allowing attachment of compatible and non-harmful devices. To allow such attachment, he proposes that industry or the FCC should define a basic air interface for wireless devices. Wu also argues for the application of net neutrality to wireless networks, which at the time meant application of the FCC's Internet Policy Statement (FCC, 2005), and stated that carriers should meter and charge for bandwidth usage rather than block particular applications. Wu also argued for wireless carrier disclosure of limits, including locks, protocol or application disabling, and bandwidth limits. Finally, he recommended that carriers and equipment manufacturers should work towards standardization of application development platforms.

In response, Hahn et al. (2007) claim that attachment of devices and Quality-of-Service (QoS) are separate issues. Having previously opposed net neutrality as a method to regulate QoS (Hahn and Litan, 2007), in this paper they argue against many of Wu's proposals. First they argue that there is sufficient wireless competition to avoid market failure and that innovation in wireless devices and applications is thriving. Next they provide an economic analysis and argue that the results show that the benefits of device subsidies, device exclusivity, and limits on devices and on applications outweigh the costs of

each.

Both of these papers focus primarily on the device attachment issue. Neither focuses on the differences in traffic management between wired and wireless networks, and hence on potential differences with respect to QoS.

The debate has entered the public policy arena. The FCC, in its *Notice of Proposed Rulemaking on Open Internet Practices* (FCC, 2009) (hereafter referred to as the NPRM), asks whether net neutrality should be applied differently to wireless networks than to wired ones. It notes that wireless networks face special challenges due to attenuation and interference, and that they determine how users and devices share scarce resources through control over the frequency, time, and power of wireless devices' transmissions. It asks whether wireless devices and/or wireless networks merit different treatment.

With respect to device attachment, the NPRM asks whether subscribers should be able to connect wireless devices of their choice providing that they do not "harm the network". It ponders whether wireless ISPs should allow attachment of any device with a compatible air interface, including tethering, and if so how to prevent harm.

With respect to services, the NPRM considers which applications or services should be covered by a net neutrality requirement. It considers whether to exclude *managed services*, defined as IP-based offerings such as voice and subscription video provided over the same networks used for broadband Internet access. It proposes that a nondiscrimination principle should only apply to Internet services, and thus exclude voice and short messaging services (SMS). However, it ponders what to do in 4th generation (4G) wireless networks capable of supporting voice, video, and data services on a converged platform architecture.

With respect to traffic management, the NPRM notes that wireless capacity may be more limited than wired capacity, and that demands can vary dynamically and widely among users. It discusses that wireless networks must be designed to deal with wide variations in signal levels across the service area as well as interference from other devices. It ponders whether wireless networks are more sensitive to user behavior. It asks how these differences in technical characteristics affect the reasonableness of various network management practices, e.g. whether it is reasonable for a wireless ISP to block particular capacity hungry applications. It also asks what impact tethering will have on wireless network congestion, and what network management measures are reasonable in this context.

The key question that this paper attempts to address is whether wireless net neutrality should be different than wired net neutrality because of the different technologies used. We recognize that readers of this paper will have a diverse

set of opinions on net neutrality. However, we expect that, independent of one's position on net neutrality, readers may wonder whether wireless networks should be treated similarly or differently than wired networks with respect to net neutrality. That is the central issue of this paper.

The first portion of the paper addresses this question without taking a position on net neutrality. However, a few basic hypotheses are necessary to focus the analysis. First, the paper focuses on whether some form of net neutrality is required to ensure a level playing field between application providers who also serve as ISPs and application providers who do not serve as ISPs; other rationales for net neutrality are not considered here. Second, it is assumed that the primary method of potential discrimination is the use of QoS mechanisms such as packet prioritization or bandwidth reservation.

The primary focus is thus placed on applications and traffic management, rather than device attachment. We are concerned with which applications or services should be covered by a net neutrality requirement, and whether this requires the definition of *managed services*. We are concerned with whether the challenges of wireless signals and mobility merit different traffic management techniques, and how these techniques may affect net neutrality.

There seem to be several parts of this issue. First, whatever one's position on net neutrality in wired networks, do the differences in the definition of *reasonable network management* or *managed services* between wired and wireless networks merit *different treatment* with respect to net neutrality? The resolution of this issue seems to depend in part on whether wireless networks require qualitatively different *types of network management* than wired networks. Second, should wireless network operators have a different ability to *restrict applications* used on their networks than wired ISPs? The resolution of this issue seems to depend in part on whether applications have a greater ability to negatively interfere with desired network operation in wireless networks than in wired networks. Third, should wireless network operators have a different ability to *restrict devices* used on their networks than wired ISPs? The resolution of this issue seems to depend in part on whether wireless devices have a greater ability to negatively interfere with desired network operation than their wired counterparts.

The focus in this paper is on the first two issues (reasonable network management and restricting applications). Section 2 reviews the pertinent aspects of network architecture and discusses the main technical differences between wired and wireless networks. Section 3 more specifically addresses how wired and wireless networks differ with respect to traffic management. With that basic understanding of network architecture, in section 4 we turn to the question of how the differences in traffic management affect net neutrality. Finally in section 5 we state our general position on net neutrality, and examine how
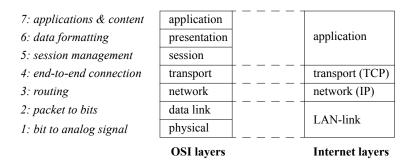
7: applications & content	application		application
6: data formatting	presentation		
5: session management	session		
4: end-to-end connection	transport		transport (TCP)
3: routing	network		network (IP)
2: packet to bits	data link		LAN-link
1: bit to analog signal	physical		
	OSI layers		**Internet layers**

Fig. 1. OSI and Internet layered models

it applies to wireless net neutrality.

2   How are wireless networks different than wired networks?

The reference models for layered network architectures are the OSI model and the Internet layering model (Braden, 1989), both pictured in figure 1.

We will consider differences between wireless and wired networks by the layer, starting at the bottom. The design of the OSI layer 1 protocol of a network is very tightly connected to the type of transmission medium used in the network, e.g. wireless, fiber, coaxial copper, etc.. Each transmission medium has different characteristics of how signals propagate through or down the medium. There are two key physical layer challenges.

First, signals become weaker (or *attenuate*) as they propagate. Wireless signals usually propagate in free space, and hence the signal energy is spread out over all directions in three dimensional space, quickly growing weaker. In contrast, wired signals propagate down a guided path, and hence the signal energy does not spread out as much. Attenuation is thus usually a greater challenge in wireless networks. Second, noise and interference cause difficulties. The environment adds background noise to the signal. In addition, a wireless signal bounces off of many objects (e.g. buildings), and these additional copies of the signal (called *multipath*) can be either beneficial or detrimental. If a wireless user is also mobile, then the amount of attenuation and multipath are continually changing.

The OSI layer 1 protocol of a network is thus matched to the characteristics of the transmission medium used in the network; different mediums warrant different LAN-link layer protocols. Signals are encoded (or *modulated*) in a manner that is effective given the attenuation and noise challenges. Wireless networks often use more complex modulation schemes than wired networks, in order to combat more severe attenuation and interference. In addition, the modulation scheme determines in what dimension users share network capac-

ity[1]. It may specify that a transmission use all of the available frequency or only a specified portion (or *channel*) of the available frequency. Many wireless modulation schemes intentionally encourage multiple transmissions to talk over each other by simultaneously using all available frequency, and the scheme sorts it out later. Finally, wireless modulation schemes continually monitor and modify the transmission power levels of each user (i.e. telling each network device to talk more softly or more loudly) to deal with changing attenuation and multipath levels.

We now move up to OSI layer 2, the data link layer. The design of the OSI layer 2 protocol of a network is also tightly connected to the type of network. The data link layer protocol, given the modulation scheme implemented in OSI layer 1, decides in real-time which users can transmit when and on which channel. Some data link layer protocols (such as those used in Ethernet and cable modems) take a laissez-faire approach. They let users attempt to transmit whenever they don't hear any other user talking. Occasionally multiple users will attempt to transmit at the same time; upon learning of the resulting packet collision, these users will have to attempt their transmissions again later. Other data link layer protocols (such as those used in DSL and telephone systems) take a more organized approach. They make users take turns and/or they assign different users to different channels. A few data link layer protocols (such as those used in 802.11 and many cell phone networks) adopt an in-between approach by giving users more direction than the laissez-faire ones but less direction than the organized ones.

Anytime that the modulation scheme and the data link layer protocol allow multiple signals to overlap in time, space, and frequency, this phenomenon is called *interference*. In wired networks, interference arises when multiple users or devices *on the same wire* transmit at the same time on the same frequency. In wireless networks, interference is caused when multiple users or devices *within hearing distance of each other* transmit at the same time on the same frequency. Interference in wireless networks is usually worse than in wired networks, because there are typically more wireless devices within hearing distance of each other than wired devices share a wire. As a result, the data link layer protocol is also carefully matched to the type of network, including whether it is wired or wireless.

We now move up to OSI layer 3, the network layer. Portions of the design of OSI layer 3 protocol of a network may depend on whether the network is wired or wireless. In both types of networks, the network layer specifies how devices addresses (e.g. IP addresses) are interpreted, and determines how a packet is routed from one network device to the next. Wireless networks, however, have additional tasks at this layer. They often must consider the allocation

---

[1] Higher layers determine how much capacity each user or application is allocated.

of wireless network resources amongst cells (part of *Radio Resource Management*), and this function is implemented in the network layer. In addition, wireless networks with mobile users must accommodate users who move from one cell to another in the middle of a call or connection (called *handoffs*). In wired networks and in some wireless networks (including Wi-Fi), the Internet Protocol (IP) and its associated routing algorithms are the dominant protocols at the network layer. In some wireless networks (including some cellular networks), a different set of protocols (Signaling System 7) emanating from telephone networks are used to implement a different method of addressing and routing. Sometimes the two sets of protocols are used together, so that higher layer Internet protocols can interface via IP to lower layer Signaling System 7 protocols.

Finally we consider OSI layers 4 through 7. These layers are responsible for managing calls or connections, responding to congestion in the core network, authenticating users, presenting data, and interfacing to applications. Although wireless networks have some special requirements at these layers, most of this functionality is common between wired and wireless networks. There are two dominant sets of protocols used. Most wired and some wireless networks use the Internet protocols, including the Transmission Control Protocol (TCP). Some wireless networks use the telephone network protocols (Signaling System 7). Sometimes variants of Internet protocols are created for wireless implementation, e.g. variants of TCP to cope with the nature of wireless attenuation and multipath.

In conclusion, wireless networks differ substantially from wired networks at the network layer and below. However, they differ in much more limited manners at transport layer and above. In the next section, we turn more specifically to how wired and wireless networks differ with respect to traffic management.

3   How is traffic management in wireless networks different than in wired networks?

As discussed in section 1, we assume that the primary concern of net neutrality is whether there exists a level playing field between application providers who are also ISPs and those who are not, and that the primary method of potential discrimination is the use of QoS mechanisms such as packet prioritization or bandwidth reservation. The question thus becomes: do the traffic management requirements of wireless networks vis a vis wired networks merit different treatment with respect to net neutrality?

The traffic management requirements of wireless networks can differ from those of wired networks for two broad reasons. First, as discussed in the pre-

vious section, wireless networks face greater challenges due to the nature of the wireless medium and often due to mobility of their users. These challenges include attenuation, multipath, interference, and handoffs.

Second, many wireless networks (especially cell phone networks) rely heavily upon their ability to offer satisfactory performance for telephone calls. Networks that specialize in supporting real-time applications allow only very limited queuing in network devices, so that each device adds only very small queuing delays. In addition, these networks reserve capacity for each call or connection or otherwise limit the total traffic in the network, so that the negative impact of users upon each other is controlled and limited.

In wireless networks, capacity reservation for voice calls is accomplished using a few different techniques. First, the power used to transmit the signal from a consumers device and vice-versa is adjusted frequently (often many times per second) on the basis of attenuation and multipath. The goal of these adjustments is to maintain a constant quality in the connection. Second, the network limits the number of number of voice users in each cell in order to maintain a minimum performance level per user. When a user migrates from one cell to another, the network attempts to allocate resources in the next cell; if resources are not available in the next cell and if the user requires too much transmission power to maintain a connection with the current cell (thus causing too much interference for its neighbors), the call is terminated. None of these techniques are yet common in the wired Internet.

These wireless traffic management techniques were designed for the dominant application on cell phone networks - telephone calls. However, there is a long term trend toward technology convergence. Both texting and Internet access are now key applications supported by cell phone networks, and subscribers pay significant sums for each. Cell phone networks are thus migrating in their architecture to look more like the Internet. In parallel, on the Internet real-time applications such as telephone calls, video conferencing, and gaming are growing in popularity. In response, Internet architecture is expected to migrate in its architecture to more efficiently support these applications. Cable tv networks now support both Internet access and telephone calls. Its architecture is also migrating toward one that can gracefully support all three sets of applications.

In order to support texting and Internet access on cell phone networks, and to create wireless local area network protocols such as Wi-Fi, traffic management techniques were required that are appropriate for delay-insensitive applications on wireless networks. These techniques have some commonality with traffic management techniques used in wired networks for delay-insensitive applications, but they must also cope with the variability of the wireless medium. Thus, they also borrow ideas from traffic management techniques

used in wireless networks for real-time applications. Like their wired network brethren, wireless traffic management for delay-insensitive applications allows for a wide variety of throughput per connection, so that a wide variety of applications can be efficiently supported. This is often accomplished using scheduling in the data link layer protocol (e.g. determining who can transmit when). Like their wireless real-time application brethren, wireless traffic management for delay-insensitive applications dynamically adjusts transmission power on the basis of attenuation and multipath, and may use some elements of Radio Resource Management to balance load between cells and to support handoffs.

It is thus true that wireless networks require some different types of traffic management than wired networks – due to both the nature of the wireless medium and the greater focus of some wireless networks on real-time applications. In wireless networks, users compete for resources with other users within hearing range, rather than only on the same wire. The greater variability of signals transmitted wirelessly instead of on wires would render real-time applications useless without some type of QoS implemented to smooth out the variations. A portion of these QoS techniques must be applied in very quick response to variations in the wireless signal, and thus must be automated in the wireless device and in network equipment.

In addition, wireless network capacity is usually more expensive than wired network capacity. The cost of wired networks is usually dominated by the cost of purchasing and installing transmission lines, and to a lesser extent by the cost of network devices such as routers. Wireless networks incur similar costs, with the cost of transmission lines replaced by the cost of obtaining spectrum. In addition, wireless networks may incur costs for leasing sites for base stations. However, the capacity of wireless networks is almost always significantly less than the capacity of wired networks that cover the same geography, due to the relative scarcity of spectrum. This decreased capacity translates into a higher shadow cost for bandwidth [2]. As a result, there is often an incentive for traffic management techniques in wireless networks to be more efficient than in wired networks.

In the next section, we turn more specifically to how these differences impact the issue of net neutrality.

---

[2] Note, however, that in low density areas wireless networks may be lower cost than wired networks, due to the ability of a single wireless base station to replace a large number of long transmission lines.

9

4 How do the differences in traffic management affect net neutrality?

Some of the questions related to net neutrality take on additional importance in wireless networks. Do users have the right to run any software of their choice on wireless devices? Do they have the right to attach wireless devices of their choice to the network?

Most proponents of wireless net neutrality believe that users should have the right to attach any device and to run any application, so long as they do not cause harmful interference to other wireless users. Most opponents of wireless net neutrality believe that an ISP can reasonably dictate which wireless devices can be used on their networks, and that either an ISP or a device manufacturer can reasonably limit which applications can be run on wireless devices; they believe that there is sufficient competition to ensure that social welfare is maximized.

In the previous section, we found that wireless networks require some different types of traffic management than wired networks. These differences included: quick response to wireless signal variation, limits on the interference users cause each other, limits on the number of active real-time users, scheduling of transmissions, and reservation or priority of resources for real-time applications. In addition, wireless network capacity is usually more expensive than wired network capacity. The question is: do these differences in traffic management justify differences in net neutrality?

For proponents, these differences pose a conundrum. They have difficulty defining *harmful interference*, and in pinpointing which traffic management practices are acceptable ways for a wireless network provider to limit harmful interference. More fundamentally, they understand that wireless telephone calls require these stronger types of traffic management, and that this includes some type of bandwidth reservation or priority. However, they wish to apply net neutrality to the remainder of the applications used. As a consequence, most proponents are left with no choice other than to somehow segment off wireless telephone calls from other applications, and to apply net neutrality only to the other applications. One such segmentation is to apply wireless net neutrality only to the *Internet* portion of a provider's offerings. Another segmentation is to apply wireless net neutrality to everything, but then to make an exception for *managed services*. Both approaches face substantial difficulties as technology convergence erases the difference between telephone networks and the Internet. How do you define the *Internet* portion when all applications, including voice and video, run over IP? Where do you draw the line between *managed services* and unmanaged services when there is a wide variety of applications that require a wide variety of QoS?

For opponents, the differences between wired and wireless traffic management are less worrisome, since the stronger techniques used in wireless seem to support their case more strongly. However, opponents have difficulty explaining why wireless traffic management requires limits on devices and applications, when such limits on wired networks would be considered unacceptable. If there is greater competition amongst wireless providers, why doesn't that translate into greater freedom, not less? Opponents often argue that the limits are justified by wireless interference. However, in both wired and wireless networks users compete with each other for resources; why does wireless interference justify limits on applications that wired interference doesn't? Alternatively, opponents argue that the limits are justified by the higher cost of wireless resources. However basic economics would lead one to expect that higher resource costs should translate into higher (or differentiated) service costs, not to somewhat arbitrary limits on use.

Thus, for both proponents and opponents of net neutrality, the differences between traffic management in wired and wireless networks pose a challenge to their positions. Having concluded in the previous section that wireless networks require some different types of traffic management than wired networks, we turn to the question of how the differences in traffic management affect net neutrality. Clearly, despite the differences in traffic management between wired and wireless networks, similar net neutrality concerns apply.

A key finding in our analysis above is that wireless networks differ substantially from wired networks at the network layer and below, and subsequently the differences between wired and wireless traffic management techniques lie almost entirely at or below the network layer, namely at OSI layers 1-3. Protocols that respond to wireless signal variation, that limit interference, that limit active real-time users, that schedule transmissions, and that reserve or prioritize resources, are all in these layers. This is consistent with the Internet layered model, in which these layers are assumed to be carefully matched to the particular transmission medium.

As a consequence, we conclude that wireless networks often are justified in using different traffic management techniques, but only at these lower layers. It follows that these differences merit a definition of *reasonable network management* that recognizes the differences in lower layer traffic management requirements and techniques. However, it also follows that lack of differences in upper layers merits a definition of *reasonable network management* that enables similar competition at the application layer independent of the type of network. Wireless networks should address their greater challenges at lower layers either by exercising stronger lower layer traffic management techniques than used by wired networks or by exercising reasonable techniques used in wired networks but to a greater extent. However, wireless networks are not justified by technical differences in implementing different traffic management

practices above the network layer than those used in wired networks.

We conclude that any net neutrality position applied to wireless networks, whether pro or con, should reflect the differences between wired and wireless networks at or below the network layer, and should reflect the similarities between wired and wireless networks above the network layer. Proponents of wireless net neutrality should accept that wireless networks require stronger forms of traffic management at or below the network layer, and should focus on regulation that ensures a level playing field between providers of various applications (including those that require QoS). Opponents of wireless net neutrality should accept that stronger traffic management at or below the network layer does not justify different treatment above the network layer, and should focus on defining reasonable network management in a manner that acknowledges this. Both sides should accept that segmentation of network applications between managed and unmanaged services, or between Internet and non-Internet offerings, is problematic as technical convergence nullifies such distinctions.

We conclude that wireless networks often are justified in using different traffic management techniques, but only at these lower layers. However, since such lower layer techniques include the reservation and prioritization methods that generated much of the initial net neutrality debate, it remains a challenge to construct a net neutrality policy that can be consistently applied to both wired and wireless networks. We turn to this challenge in the remaining section of the paper.

5 Wireless net neutrality versus wired net neutrality

Up to this point in the paper, we have addressed the differences between wired and wireless net neutrality without taking a position on the issue of net neutrality itself. Now we consider potential methods to extend wired net neutrality to wireless networks, and in doing so we do take a position. In this section, we give our general position on net neutrality, and consider how it may apply to wireless net neutrality.

In Jordan (2007), we argued for middle-ground on net neutrality. We proposed a policy that (i) bans discrimination on the basis of source, destination, or ownership of traffic, (ii) allows ISPs to implement other QoS mechanisms, (iii) requires that any such QoS mechanisms implemented in network infrastructure be made available without unreasonable discrimination to competing application providers and peering ISPs, (iv) allows ISPs to charge subscribers, application providers, and/or peering ISPs for use of such QoS mechanisms provided these charges are not unreasonably discriminatory, (v) bans unfair

methods of competition and unreasonable discrimination, and (vi) provides forbearance when sufficient competition exists in network infrastructure.

The policy is based on a framework that distinguishes between discrimination in high barrier-to-entry network infrastructure and in low barrier-to-entry applications. The policy prohibits use of Internet infrastructure to produce an uneven playing field in Internet applications. In this manner, the policy restricts an Internet service provider's ability to discriminate in a manner that extracts oligopoly rents, while simultaneously ensuring that ISPs can use desirable forms of network management. The paper presented statute language to implement the proposed policy.

On a related note, in Jordan (2009), we extended our analysis to consider other types of traffic management that might fall under the umbrella of net neutrality. We proposed that traffic management practices are reasonable if they are implemented at endpoints, are chosen by the user, are based on reasonable application provider payment, or involve providing QoS to traffic chosen by the user. We proposed that traffic management practices implemented in transit nodes without user choice are unreasonable if they block applications or violate the net neutrality policy discussed above (e.g. provide QoS based on source or on unreasonably discriminatory application provider payment). We suggest that QoS based on application can be more effectively implemented by allowing the user to determine the priority of his/her applications, and we suggest that any charges for QoS can be most effectively implemented by integrating them into subscriber contracts and into the Service Level Agreements between ISPs, rather than by charging application providers that are not subscribers.

We acknowledge portions of this middle-ground position may be unacceptable to some net neutrality proponents and some opponents. Some proponents of net neutrality object to allowing ISPs to implement QoS or to allowing them to charge for it. Some opponents of net neutrality object to prohibitions on deep packet inspection or on unreasonably discriminatory charges.

We now turn to the question of whether such policies on wired net neutrality can be extended to wireless net neutrality. In both wired and wireless networks, unreasonably discriminatory use of reservation of bandwidth and/or prioritization of packets can threaten to tilt the playing field between ISPs and application providers or amongst application providers. The immediate danger in wired networks is that cable ISPs who prioritize their own VoIP traffic may deny access to this same prioritization to competing VoIP providers or may ask unreasonably prices for this QoS, and that ILECs who prioritize their own video over IP traffic may deny access to this same prioritization to competing video providers or may ask unreasonably prices for this QoS. We see the same danger in wireless networks, perhaps even to a greater extent

due to the more widespread use of QoS.

The question is how to limit these dangers. One option is to ban the use of reservation and prioritization practices. This option is unreasonable, since these QoS practices are often necessary to support real-time applications. A similar option is to restrict their use to *managed services*. This option is impractical, since as discussed above we don't believe one can define such a class; applications change very rapidly, and it is this flexibility and evolution that must be maintained. Rather than banning or limiting QoS, some proponents of net neutrality would mandate open access to all lower layer components and protocols to application providers. We don't believe this is necessary to accomplish the goals.

Requiring an *open interface* to QoS mechanisms in lower layers is simpler and more efficient than requiring open access or defining *managed services*. The idea comes from separating networking functions into two subsets: OSI layers 1-3 (network infrastructure) and OSI layers 4-7 (network applications). Network infrastructure layers exhibit a high barrier-to-entry and hence have small number of providers, while the application layers exhibit a low barrier-to-entry and hence have a large number of providers. Net neutrality can be effectively achieved by a properly regulated open interface *from applications to network infrastructure*. The requirement of an open interface captures the central tenet of a layered Internet architecture, and yet is less intrusive than a requirement of open access. In particular, it is less intrusive than requiring a standard air interface for wireless devices, which may be required to ensure the right to attach devices; an open interface between the networking and transport layer only requires standardization of the *Application Programming Interface* (API) for services offered by the networking and lower layers, while a standardized air interface requires standardization of all physical and data link layer protocols and of their use.

In Jordan (2007), we proposed that any QoS mechanisms that an ISP implements in network infrastructure layers should be available to application providers without unreasonable discrimination. Requiring such an open interface can ensure that ISPs are prohibited from refusing to provide enabling Internet infrastructure services to competing application providers in order to differentiate the ISP's own application offerings, prohibited from providing Internet infrastructure services to competing application providers at inflated prices in order to favor the ISP's own application offerings, and prohibited from making exclusive deals to provide enabling Internet infrastructure services to certain application providers. It can also ensure that ISPs have the right to apply network management mechanisms that do not threaten a level playing field, and to make arrangements with consumers, application providers, and peering ISPs for Internet infrastructure services in a manner that does not conflict with the above goals. Finally, forbearance should be applied to these

regulations where sufficient competition exists.

But is an open interface an effective way to require wireless net neutrality? The differences between wired and wireless traffic management lie almost entirely within the network infrastructure layers. Hence, an open interface would similarly require that the QoS mechanisms used in these wireless layers be available without unreasonable discrimination to application providers. There is no need to define *harmful interference*, as the ways in which users compete for resources with each other continues to be controlled by the network infrastructure layers; an open interface does not impede an ISP's ability to reasonably accomplish this. There is no need to define what constitutes the *Internet* portion of a provider's offerings, as the layering model applies both to Internet protocols and to telephone network protocols; all of a wireless providers offerings are thus included without reference to the application or technology. There is also no need to define what constitutes *managed services*, as the open interface requires access to lower layer QoS mechanisms that enables real-time applications, rather than carving out real-time applications as an exception; the open interface thus encourages competition in managed services, rather than inhibiting such competition. Finally, requiring an open interface in both wired and wireless networks avoids the need to differentiate net neutrality policy on the basis of the technology used. However, it does allow for differences between wired and wireless networks on the basis of competitiveness, since forbearance may be granted in some markets but not in others.

An open interface also would bear on the question of whether users have the right to run the software of their choice on wireless devices. Because an open interface allows any application provider to offer any functionality within the network application layers without unreasonably impacting the network infrastructure layers, there is no reasonable justification on the basis of traffic congestion for limiting applications on any device. The impact of an application would be controlled by the ISP at the interface through limits and/or charges placed on traffic. Applications that transmit high volumes of traffic would either purchase this capacity at standard rates or would consume a high proportion of a pre-purchased traffic quota. Applications that require QoS would similarly either purchase the required QoS at standard rates or would consume some portion of pre-purchased QoS. The interface is only concerned with traffic volume, QoS, and payment; it is not concerned with what particular application this traffic is destined to. From the point of view of the lower layers, therefore, the system is application-agnostic.

This interface-based approach is consistent with current wireless device and operating system architecture. There are a number of different operating systems that are used on wireless devices. However, they are all built using layered architectures. The operating system limits access of applications to lower layer

protocols. Indeed, they provide an interface between applications and the operating system called an *Application Programming Interface* (API). The API defines what services the operating system offers to applications and how to access them. As in wired networks, common wireless network device operating system APIs offer access to lower layer functionalities at the the networking layer or above; they do not offer direct access to protocols at the physical and data link layers. This architecture implies that *open access* to all lower layer protocols is not necessary for net neutrality; all that is required is an *open interface*. Similarly, the architecture implies that lower layer protocols have no need to know what application a packet belongs to; they only need know what type of traffic management to apply.

6 Conclusion

The principal question addressed in this paper is whether differences between wired and wireless network technology merit different treatment with respect to net neutrality. Unlike other papers in the academic literature, we focus on applications and traffic management, rather than device attachment. We found that wireless networks differ substantially from wired networks at the network layer and below, and that wireless networks often require different traffic management practices at these lower layers. However, since the differences are confined to these lower layers, we argued that net neutrality in both wired and wireless networks can be effectively accomplished by requiring an open interface between network and transport layers.

7 Acknowledgment

References

Braden, R., 1989. Requirements for Internet hosts – communication layers. Tech. rep., IETF RFC 1122.

Clark, D. D., 2007. Network neutrality: Words of power and 800-pound gorillas. International Journal of Communication 1, 701–708.

FCC, 1968. FCC 68-661, Carterfone Order.

FCC, 2005. FCC 05-151, Internet Policy Statement. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-05-151A1.pdf.

FCC, October 2009. FCC 09-93, Open Internet NPRM. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-09-93A1.pdf.

Hahn, R. W., Litan, R. E., 2007. The myth of network neutrality and what we should do about it. International Journal of Communication 1, 595–606.

Hahn, R. W., Litan, R. E., Singer, H. J., 2007. The economics of "wireless net neutrality". Journal of competition law and economics 3, 399–451.

Jordan, S., 2007. A layered network approach to net neutrality. International Journal of Communication 1, 427–460.

Jordan, S., August 2009. Some traffic management practices are unreasonable. In: Proceedings of the IEEE International Conference on Computer Communications and Networks (ICCCN). pp. 137–140.

Peha, J. M., 2007. The benefits and risks of mandating network neutrality, and the quest for a balanced policy. International Journal of Communication 1, 644–668.

Wu, T., 2007. Wireless Carterfone. International Journal of Communication 1, 389–426.

How to determine whether a traffic management practice is reasonable

Scott Jordan

Department of Computer Science, University of California, Irvine

Arijit Ghosh

Department of Computer Science, University of California, Irvine

Abstract

As part of the wider debate over net neutrality, traffic management practices of Internet Service Providers have become an issue of public concern. An increasing number of vendors offer network equipment to implement a variety of traffic management practices using packet classification and packet filtering. The Federal Communications Commission has asked for public input on whether such traffic management practices are reasonable forms of network management. Little attention has been paid to this issue within the academic computer science community, and many communications policy researchers have recommended a case-by-case analysis. In contrast, in this paper we propose a framework for the classification of traffic management practices as reasonable or unreasonable. To build the framework, we focus both on the technical aspects of traffic management techniques and on the goals and practices of an ISP that uses these techniques. The framework classifies traffic management practices as reasonable or unreasonable on the basis of the *technique* used and on the basis of *how and when* the techniques are applied. We suggest that whether a traffic management practice is reasonable largely rests on the answers to four questions regarding the techniques and practices used. We consider examples of how these techniques are used by ISPs, and how the answers to these four questions collectively affect the degree to which a traffic management practice is reasonable. Based on these questions, we propose a framework that classifies techniques as unreasonable if they are unreasonably anti-competitive, cause undue harm to consumers, or unreasonably impair free speech.

Email addresses: `sjordan@uci.edu` (Scott Jordan), `Arijit.Gjosh@uci.edu` (Arijit Ghosh).

1 Introduction

The traffic management practices of Internet Service Providers (ISPs) have become an issue of public debate. In 2007, Comcast started using reset packets to terminate selected peer-to-peer connections (Comcast Corporation, 2008). This practice, when uncovered by a few users, generated a firestorm of debate, largely because it dovetailed into an existing debate over net neutrality (Weitzner, 2008).

Net neutrality represents the idea that Internet users are entitled to service that does not discriminate on the basis of source, destination, or ownership of Internet traffic. Proponents of net neutrality argue that without a prohibition on discrimination, ISPs may charge application providers discriminatory prices for access to dedicated bandwidth or for quality of service (QoS), or may outright block access to certain applications or websites, and that such activity will inhibit development of new Internet applications (Jordan, 2009). To proponents of net neutrality, Comcast's practices seemed like blocking of certain applications; to Comcast, however, its practices seemed like reasonable traffic management designed to limit network congestion.

The debate centers not only on Comcast's practices, but also on the wider use of deep packet inspection techniques which allow ISPs to identify and control traffic streams on the basis of transport and application layer information. An increasing number of vendors offer equipment that can be placed in the network to implement a variety of traffic management practices using packet classification and packet filtering.

In response to the early net neutrality debate in the United States, in 2005 the Federal Communications Commission (FCC) issued a set of principles (FCC, 2005b). The principles express the sentiment that consumers should be entitled to connect devices and to access content and applications of their choice. In a footnote, the FCC comments that these principles are subject to "reasonable network management", but does not define what this terms means. In response to the discovery of Comcast's traffic management practices, a few organizations petitioned the FCC to rule that an ISP is violating these principles (and thus not practicing reasonable network management) when it intentionally degrades a targeted Internet application such as peer-to-peer (Free Press, Public Knowledge et. al., 2007) and to adopt rules that would prevent such practices (Vuze Inc., 2007).

The FCC asked for public input on whether this practice and other traffic management practices are reasonable forms of network management (FCC, 2007). They asked whether ISPs use traffic management practices to prioritize latency-sensitive applications, to block unwanted traffic, to implement

parental controls, to improve network performance, or to gain advantage over competitors. They also asked whether these practices are helpful or harmful to consumers and whether they are reasonable. In 2008, the FCC concluded that Comcast violated a principle concerning users' rights to access lawful Internet content and use applications of their choice, and that its practices do not constitute reasonable network management (FCC, 2008). However, the FCC has not adopted rules that would delineate reasonable network management.

These questions have largely gone unanswered by the academic networking community. Most networking technologists would have some concern about violations of layering such as that involved in deep packet inspection. However, there is no consensus about when layering violations are warranted or how to respond to them.

There have been only a few attempts in the networking literature to go beyond the technical aspects of traffic management and to consider the social and legal implications. Weitzner (2008) discusses the Comcast incident and the connections to net neutrality. Peha (2007) discusses the incentives that ISPs may have for using discriminatory practices, and the benefits and damages that may accrue from these practices. He also gives examples of what should be allowed and prohibited, but does not give a framework that allows one to classify practices. Frieden (2006) similarly gives examples of what he believes to be permissible and impermissible traffic management practices, and suggests a few best practices (including limitations on blocking and degradation) that ISPs should adopt. He similarly does not present a framework for classification, but instead proposes that the FCC should impose reporting requirements on ISPs and assess practices on a case by case basis. Lehr et al. (2007) discuss strategies that end-users may adopt in response to ISP discrimination, including technical counter-measures.

However, we have found no literature that proposes a method for classification of traffic management practices as reasonable or unreasonable. In this paper, we present such a framework for traffic management by Internet Service Providers within the United States. We restrict our attention to traffic management policies as a subset of a larger class of network management policies. To build the framework, we focus both on the technical aspects of traffic management techniques and on the goals and practices of an ISP that uses these techniques. The framework classifies traffic management practices as reasonable or unreasonable on the basis of the *technique* used and on the basis of *who* decides when the techniques are applied. The framework results in classifying practices as unreasonable when they are unreasonably anti-competitive, cause undue harm to consumers, or unreasonably impair free speech.

The paper proceeds as follows. In section 2, we suggest that whether a traffic management practice is reasonable largely rests on the answers to four

questions regarding the techniques and practices used. Section 3 considers examples of how these techniques are used by ISPs, and how the answers to these four questions collectively affect the degree to which a traffic management practice is reasonable. Based on these questions, in section 4 we propose a framework that classifies techniques as unreasonable if they are unreasonably anti-competitive, cause undue harm to consumers, or unreasonably impair free speech.

2   Key questions about traffic management techniques and practices

Traffic management is applied to implement a variety of functions, at a variety of layers, by a variety of actors, in a variety of manners, for a variety of purposes. To delineate these components, define a *traffic management technique* as a specific function that is offered at a specific layer. The function should determine whether traffic is transmitted and/or the rate at which traffic is transmitted, or should enable such functions in other techniques. Define a *traffic management practice* as a collection of traffic management techniques, used by a specific type of actor, in a specific manner, for a specific purpose. In this section we suggest that whether a traffic management practice is reasonable largely rests on the answers to four questions regarding the techniques and practices used. In the next section, we will consider examples of how these techniques are used by ISPs to form traffic management practices.

The first two questions apply to traffic management *techniques*, because they are directed at the layer ("where") and functionality ("what"). The second two questions apply to traffic management *practices*, because they are directed at the actor ("who") and the manner and purpose ("when").

The first question is:

(1) WHERE: Where in the network, and at which layer, is the traffic management technique applied?

To answer the "where" question, we need to understand layers. The Internet is based on the concept of a layered architecture, where each layer provides certain functionalities. The reference model for layered architectures is the OSI model, developed by the International Standards Organization. The OSI model is composed of 7 layers, while the Internet model collapses this down to 4 layers, as pictured in figure 1. It is useful to think of the physical connection, e.g., wire, as being located below the bottom-most layer (layer 1) and the user, e.g., you, as being located above the top-most layer (OSI layer 7). Traffic management can be applied in any of the layer. Access control is often im-
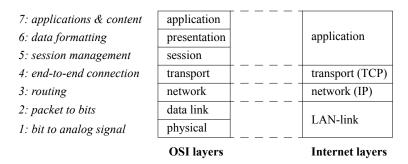
4

7: applications & content	application		
6: data formatting	presentation		application
5: session management	session		
4: end-to-end connection	transport		transport (TCP)
3: routing	network		network (IP)
2: packet to bits	data link		LAN-link
1: bit to analog signal	physical		
	OSI layers		**Internet layers**

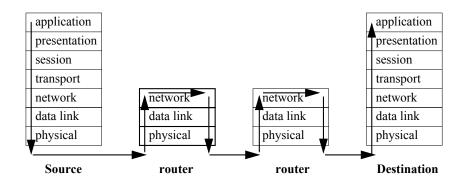Fig. 1. OSI and Internet layered models

Fig. 2. Routing

plemented in the Internet Transport or Application layers, e.g. the Integrated Services (IntServ) protocol is at the transport layer. Traffic shaping can be implemented at any Internet layer, e.g. the TCP protocol at the Transport layer, the Differentiated Services (diffServ) protocol at the Network layer, and similar protocols which control flow on a local area network at the LAN-link layer. Techniques that use deep packet inspection (DPI) are usually operating at multiple layers, e.g. termination of selected peer-to-peer connections identified using DPI operates at both Transport and Application layers.

It is worthwhile to distinguish between techniques operating purely at endpoints versus techniques operating at transit points such as routers. Not every network device contains all 7 layers. Personal computers do contain all 7 layers - the network interface card (e.g., Ethernet card) implements OSI layers 1 and 2, the operating system (e.g., Windows) implements OSI layer 3 and part of layers 4 through 7, and user-installed software implements the remainder of layers 4 through 7. A network router, however, often contains only layers 1 through 3. As a result, communication from source to destination follows a complicated path as pictured in figure 2. Indeed, one of the Internet's key design paradigms, the end-to-end principle (Saltzer et al., 1984) suggests that network functionality should be implemented in OSI layers 1 through 3, and hence in each router, only if it cannot be implemented effectively in higher layers. This principle has been followed in much (but not all) of Internet design to date.

With respect to the "where" question, we thus propose that the pertinent distinction should be whether the technique is applied (a) at or above transport layer *and* in a transit node or (b) either below transport layer *or* at an endpoint. If a technique is applied at or above the transport layer, then good network design recommends that it be applied only at an endpoint. Therefore, techniques that are applied at or above the transport layer *and* in a transit node likely violate this design principle; although this violation is not sufficient to make a traffic management practice unreasonable, it should raise a red flag. None of the standardized traffic management techniques discussed in the previous section is intended to be used in a manner that violates this principle. However, a number of proprietary traffic management techniques do violate layering. Proprietary products such as Sandvine (Sandvine Incorporated, 2004) or PacketShaper (Packeteer Incorporated, 2007) are used in transit nodes and involve DPI; this violates layering since transit node devices should not inspect transport or application layer headers or the application layer payload. Firewalls also violate layering if they are placed in transit nodes.

If a technique is applied below transport layer, then the design principle allows implementation at transit nodes. For instance, guaranteed QoS can only be provided by offering QoS in every portion of the network that may experience congestion. Thus, the reasonableness of traffic management practices implemented below transport layer may also depend on whether such techniques are available at transit nodes.

Are these practices reasonable? The answer to the "where" question alone is not sufficient to make this determination; other questions must be considered.

The second question is:

(2) WHAT: What type of traffic management functionality is applied?

The purpose of a set of traffic management techniques is to management the overall performance of the network. The goal of this performance, however, must come from the objective of the network operator, which is considered below. However, one can usually define this goal either in terms of ensuring a maximum acceptable congestion level or in terms of maximizing revenue. Given a goal, the network operator has the choice of a wide variety of techniques to try to achieve this goal, and will usually apply several techniques that complement each other.

Different techniques work on different time scales. For congestion that lasts only a short time (e.g. less than a second), traffic shaping at transit points is generally considered effective. If the goal is to ensure a maximum overall acceptable congestion level, then the burst of packets is simply queued. If there are multiple types of flows, then priority service techniques such as diffServ and weighted fair queueing can ensure different minimum performance

levels for each type or can attempt to maximize revenue. If the congestion lasts for more than a short time, however, such techniques are insufficient and queues overflow. For congestion that lasts for a moderate time (e.g. more than a second but less than a minute), traffic shaping at endpoints is generally considered effective. Applications that can tolerate large fluctuations in instantaneous throughput, such as file transfer, email, or web browsing, are called *elastic*. End-to-end flow control, usually TCP, can be applied to such elastic applications, and this is typically an effective method of reducing moderate time scale congestion. If the congestion lasts for more than a moderate time, however, such techniques result in unacceptable performance for inelastic applications. Therefore, for congestion that lasts a long time (e.g. more than a minute), access control is often required to reduce the number of flows competing for resources.

Access control is thus a more severe form of traffic management than end-to-end flow control, which is in turn more severe than traffic shaping at transit points. With respect to the "what" question, we thus propose that the pertinent distinction should be whether the functionality of the traffic management technique is (i) blocking or termination of a session versus (ii) enhancement or degradation of QoS. Blocking or termination is a severe form of traffic management and should raise a red flag. In contrast, enhancement or degradation of QoS is much less severe if applied in moderation.

The technique must be matched broadly to the type of application, since a user's perception of performance depends on the way the application responds to fluctuations in QoS. The choice is generally between suffering temporary but significant reductions in rate and having your connection blocked or terminated. Users of elastic applications would generally prefer that throughput fall in times of congestion rather than the flow be blocked or terminated; thus end-to-end flow control is appropriate. In contrast, Users of applications that are inelastic, e.g. voice over IP (VoIP) or video conferencing, are unwilling to tolerate a significant reduction in rate and are upset with call terminations and would prefer that the connection be blocked before it starts if network capacity is insufficient to maintain an acceptable quality connection.

The majority of the traffic management techniques discussed above use enhancement or degradation. A few, however, use blocking or termination. The IntServ architecture includes provisions to block new connections if adequate resources are unavailable. VoIP applications may block or terminate connections if sufficient QoS cannot be maintained. Sandvine's traffic management products can terminate selected TCP connections (Sandvine Incorporated, 2007). Firewalls are intended to block selected connections. The red flag raised by the use of blocking, therefore, is not sufficient to determine whether a practice is reasonable; other questions must be considered.

The third question is:

(3) WHO: Who decides whether the traffic management practice is applied?

We propose that the pertinent distinction should be whether the traffic management practice is applied (i) directly by a user or by an ISP only when a user desires this action versus (ii) by an ISP independent of a user's wishes. A user or a user's application can communicate its desires in several manners. If the technique is applied solely at the endpoints, then the source and destination devices can jointly determine whether and how to use the technique without the need for network operator participation. If the technique requires implementation at transit points, then a user can be presented with choices by the network and can respond according to the user's objectives; most likely this would be done in an automated manner by the user's application based on a profile selected by the user. For example, if the network used IntServ to reserve network capacity for some flows, then the user could decide which flows should receive enhanced performance. Similarly, if the network used diffServ to prioritize some packets, then the user could mark those packets that should receive priority. In this manner, the user decides which applications are important to him/her and the perceived value of good performance.

Actions taken by a user or under the user's direction are generally not deemed to be unreasonable. However, actions taken unilaterally by an ISP should raise a red flag, worthy of further investigation. Examples of each of these are given in the next section.

The final question is:

(4) WHEN: On what basis is it decided to apply the traffic management practice?

Traffic management can be used in various manners and for various purposes. Rather than relying on case-by-case analysis, we propose that the pertinent distinction should be whether the traffic management practice is applied to certain traffic on the basis of (i) the application, (ii) the source and/or destination, (iii) service provider, and/or (iv) payment. Practices applied to certain applications may be reasonable if they are done in a nondiscriminatory manner. Practices applied to traffic based on source and/or destination, however, are likely to raise a red flag out of anti-competitive concerns. Similarly, practices applied only to traffic carried by certain service providers are likely to raise a red flag for the same reason. In contrast, the reasonableness of practices applied on the basis of payment is likely to rest on the reasonableness of the payment amount. Examples of each of these are given in the next section.

3   Traffic management practices

In the previous section, we proposed four questions that affect the degree to which a traffic management practice is reasonable or unreasonable. In this section, we consider five examples of traffic management practices and discuss how the answers to these four questions affect their reasonableness. In the next section, we will use the lessons learned here to construct a framework for determination of whether a traffic management practice is reasonable.

First, consider the use of session management techniques that started this debate over traffic management practices. Sandvine's traffic management products are capable of identifying and terminating file-sharing connections (Sandvine Incorporated, 2004, 2007). In 2008, Comcast used products such as Sandvine to terminate TCP connections carrying BitTorrent packets used for uploading files from a Comcast subscriber to a destination outside the Comcast network, when the Comcast subscriber was not simultaneously downloading files (Comcast Corporation, 2008). For this traffic management practice, the answers to the four questions are:

Where: at or above the transport layer, in a transit node (red flag).
What: termination (red flag).
Who: by an ISP independent of a user's wishes (red flag).
When: on the basis of the application and the destination (possible red flag).

This practice raises at least three red flags: (1) it violates layering, because a transit node operates at or above the transport layer; (2) it involves termination of a connection; and (3) it is done independent of a user's wishes. With so many red flags, we easily find this practice to be unreasonable. The principal reason is that causes undue harm to consumers, since there are more direct and transparent manners to limit traffic from a user. Indeed, the FCC concluded that the practice is unreasonable, by relying on the following aspects of the practice: blocking, anti-competitive harm, lack of disclosure, and lack of tailoring of the practice to combat network congestion (FCC, 2008).

Next, consider another class of practices that involves blocking or termination of connections – firewalls. The answers to the four questions for firewalls are:

Where: at or above the transport layer (ok), at the endpoint or in transit nodes (red flag).
What: blocking (red flag).
Who: directly by a user or by an ISP only when a user desires this action (ok), or by an ISP independent of a user's wishes (red flag).
When: on the basis of the application and/or the source and/or destination (possible red flag).

The use of firewalls as a traffic management practice can thus also raise several red flags. First, firewalls can be implemented in endpoints (e.g. Windows Firewall) or in transit nodes (e.g. in wireless routers or network gateways). When implemented in transit nodes, this is a layering violation which raises a red flag. In addition, firewalls such as parental control software can be used to block traffic from certain sources, which raises another red flag. However these uses of firewalls are universally accepted as reasonable forms of traffic management. Why? The answer is that such firewalls are under the control of the end user. In contrast, firewalls have sometimes been used by ISPs independent of a user's wishes. In 2005, Madison River Communications blocked ports used by VoIP applications, which the FCC concluded is unreasonable traffic management (FCC, 2005a). Currently, many ISPs block connections to or from specific ports to combat spam (e.g. blocking outgoing SMTP traffic to port 25) or to prohibit residential servers (e.g. blocking incoming traffic to selected server ports). While combating spam is a worthy goal, users may desire to send email via other ISP's servers; hence, we conclude that the use of firewalls in this manner is a traffic management practice that should be used only with the consent of the user. Similarly, while ISP contracts may prohibit operation of a residential server, there are more direct and transparent manners to limit traffic to and from a user, and this practice should be considered unreasonable. We conclude that if the "who" question is resolved in favor of user choice, then the other red flags do not matter.

What about traffic management practices that involve limited degradation of traffic without blocking or termination? Many products offer proprietary traffic shaping techniques, and a number of ISPs use these techniques to limit file-sharing traffic. The answers to the four questions for this practice is:

Where: at or above the transport layer, in a transit node (red flag).
What: degradation (possible red flag).
Who: by an ISP independent of a user's wishes (red flag).
When: on the basis of the application (possible red flag).

Many educational institutions implement this practice by configuring products such as PacketShaper to limit the network bandwidth used by file-sharing applications (Packeteer Incorporated, 2008). The practice delays the transmission of file-sharing packets and hence slows down the rate at which these streams are forwarded through the device. This type of traffic shaping could be implemented at the network layer if low priority packets were labeled by the user. However, without the user's involvement to identify low priority packets, products such as PacketShaper use DPI to determine which packets belong to file-sharing applications. Use of DPI classifies this practice as an application layer practice; because an application layer practice is applied at a transit node, it violates layering, which raises one red flag. A second red flag is raised because the practice is typically applied without the consent of the user. This

type of practice is less severe than blocking of termination; opinions differ as to whether these two red flags are sufficient to classify the practice as unreasonable. Since there are more direct and transparent manners to limit traffic from a user, we are reticent to classify such techniques as acceptable. However, because these alternative practices involve different business models that may require some time to be accepted by the public, we recommend classifying traffic shaping for file-sharing traffic as a borderline traffic management practice that could be used for a limited period of time if properly disclosed in the user contract.

Next, consider another class of practices that involves limited degradation – tiering. The answers to the four questions for tiering are:

Where: at or below the network layer, in a transit node (ok).
What: degradation (possible red flag).
Who: by an ISP on the basis of a user's wishes (ok).
When: on the basis of consumer payment (ok).

Tiering is typically accomplished in transit nodes (the user modem and/or ISP routers) at the data link and network layers by limiting the user download and upload rates to the maximum rates dictated in the user contract. This is a form of degradation, since the equipment is capable of transmitting at higher rates. However, since this practice is applied on the basis of user choice (and clearly displayed in user contracts), this is universally considered to be reasonable traffic management.

Finally, consider an example of a traffic management practice that involves enhancement of QoS. Currently, this is commonly used to support an ISP's own offering of VoIP or video-over-IP:

Where: at or below the network layer, in transit nodes (ok).
What: enhancement (possible red flag).
Who: by an ISP on the basis of a user's wishes (ok).
When: on the basis of the application and the service provider (possible red flag).

Enhanced QoS for real time applications such as voice and video typically requires the use of traffic management techniques that offer QoS in the data link and/or network layers in every portion of the network where congestion may occur (see e.g. Cox Communications (2004)). When an ISP uses enhanced QoS for its own VoIP and/or video-over-IP offerings, it uses these practices within its own network. In the case of VoIP, the traffic is then transited onto the public switched telephone network which offers similar QoS. In the case of video-over-IP, the video source usually resides on the ISP's network, so the entire network path (up to the subscriber premises) is under the control of the ISP. Although the practice is applied without the ability for a user to decline

this enhancement, presumably no user would desire their voice or video service to have a lower QoS. This practice does however raise one red flag because it is applied only to voice and/or video service offered directly by the ISP. We do not object to the use of QoS, nor to charging for QoS; however, we do believe it is an acceptable traffic management practice *only if* the ISP offers the same QoS service for services offered by other providers at a rate that is not unreasonably discriminatory (Jordan, 2009).

4 A framework for determination of whether a traffic management practice is reasonable

In the previous two sections, we proposed four questions that affect the degree to which a traffic management practice is reasonable or unreasonable, and investigated the reasonableness of five examples of traffic management practices on the basis of the answers to these questions. In this section, we propose a framework for determination of whether a traffic management practice is reasonable.

The order in which the questions are considered is important. Start with one part of the "where" question, the location in the network where the traffic management technique is applied. If the technique is *applied at an endpoint*, we propose that it be classified as a reasonable traffic management practice regardless of the answers to the other questions. One endpoint is the user; practices applied directly by the user are not in question. The other endpoint is the entity with which the user is communicating. When this entity is an ISP, the ISP is acting in the role of an application provider. Common examples of this situation are ISPs that offer email and/or web hosting services. However, a user can (or should be able to) receive such application services from a large number of potential providers. Since this market is competitive, practices applied at an endpoint that negatively impact the user's experience may drive users to change application providers, buy they need not change their ISP. Therefore, any traffic management practice applied at an endpoint should be classified as reasonable. In contrast, if the traffic management practice is applied *at a transit node*, we must consider the remaining questions.

Next consider the "who" question, namely who decides whether the traffic management practice is applied. If the traffic management practice is applied *directly by a user or by an ISP only when a user desires this action*, we propose that it should be classified as a reasonable traffic management practice because the user has control over whether the practice is applied. Such practices are common, and include many firewalls, parental control software, and tiering. If an ISP were to provide enhanced QoS for voice or video purely on the basis of consumer payment, then this payment for QoS would not be

discriminatory and we propose that it be classified as a reasonable traffic management practice. In contrast, if the traffic management practice is an action *taken unilaterally by an ISP*, then it is worthy of further investigation. If a practice is used without user consent, then we believe it should be disclosed in sufficient detail in the user contract. If so disclosed, then we must consider the remaining questions to determine if it is a reasonable practice.

Before progressing to these remaining questions, however, we should limit the scope of the traffic management practices considered here. We only consider techniques that are applied to networks such as the the Internet that use a public right-of-way; private networks are free of such regulation. We only consider techniques that affect Internet applications; if an ISP offers a voice service under Title II of the Communications Act (which regulates common carriers) or offers a video service under Title VI of the Communications Act (which regulates cable communications), then these restrictions need not apply. We only consider lawful uses; ISP rights to detect and interfere with illegal uses are addressed elsewhere in law. we only consider non-harmful uses of the network; security measures may require special considerations. We do not consider issues of privacy, which intersect with many of the techniques discussed here but which require considerations beyond those detailed here. Finally, prohibition of unreasonable practices should implemented only where sufficient competition does not exist; Title I of the Communications Act includes a provision which instructs the FCC to forbear from applying regulations unless they are in the public interest and required to ensure just and reasonable practices. Toward this end, regulation of reasonable traffic management should only apply to access networks, specifically to the portions of an ISP's network which must be transversed to form routes from the Internet to its subscribers.

The next aspect to be considered is the "what" question, in particular whether the practice involves blocking or termination of a session versus enhancement or degradation of QoS. If the practice involves *blocking or termination*, we propose to classify it as unreasonable. Blocking or termination practices that are applied at a transit node without user choice are unreasonably anti-competitive, cause undue harm to consumers, or unreasonably impair free speech. When blocking is applied at a transit node without user choice *on the basis of the source or destination or on the basis of the speech within the packet*, the practice unreasonably impairs free speech; this type of blocking includes blocking of specific web pages or blocking on the basis of the content of the speech. When blocking is applied at a transit node without user choice *on the basis of the application*, the practice is unreasonably anti-competitive and/or causes undue harm to consumers; this type of blocking includes blocking of specific applications (e.g. blocking or terminating VoIP or file-sharing connections) and blocking of specific ports (e.g. SMTP or server ports). There is no reasonable justification for the use of these techniques. In some cases, the ISP's goal may be to limit congestion, reduce spam, or implement security;

13

however, such goals can be implemented either through less severe methods that do not involve blocking or with the consent of the user. If a traffic management practice is implemented in a transit node, without user choice, but does not block or terminate connections, we must consider the remaining questions.

Practices that *enhance or degrade QoS in a transit node without user choice* are the concern of the remainder of this section of the paper. To address such practices, consider the "when" question, which asks on what basis is it decided to apply the traffic management practice. This question considers the manner and purpose of the practice. We propose that the pertinent distinction should be whether the traffic management practice is applied to certain traffic on the basis of (i) the application, (ii) the source and/or destination, (iii) service provider, and/or (iv) payment.

First, consider using *source and/or destination and/or service provider* as the basis. A common example of this practice is an ISP that provides enhanced QoS for its own VoIP service, but does not provide this same QoS to competitors VoIP packets. Another example of an exclusive arrangement would occur if an ISP were to provide access to enhanced or degraded QoS to some third party application providers but not others. Use of source and/or destination and/or service provider without user choice involves the use of exclusivity. Such exclusive arrangements are unreasonable, since they tilt the playing field between application providers through use of Internet infrastructure. Thus, we propose that these traffic management practices be classified as unreasonable, because they are unreasonably anti-competitive.

Next, consider using *payment* as the basis for the decision of when an ISP uses enhanced or degraded QoS. For instance, an ISP could charge a consumer for enhanced QoS for all packets to or from that subscriber. Alternatively, an ISP could charge an application provider for enhanced QoS for all packets to or from that application provider. Consumer payment for QoS places the use of the practice under the control of the user, and hence this framework would already have classified such practices as reasonable. We thus only need consider charging of application providers. We considered this case in detail in Jordan (2009). If the price is not unreasonably discriminatory (e.g. if an ISP sells QoS to all application providers at the same price as it passes on to its own applications that require QoS), then we argued in Jordan (2009) that the practice is reasonable. However, if prices for QoS are unreasonably discriminatory, then a traffic management practice that uses such prices as the basis is unreasonable since the practice is unreasonably anti-competitive.

Finally, consider cases in which the practice is applied on the basis of the *application*. In these cases, if the practice is applied entirely *at or below the network layer*, then we propose that the practice be classified as reasonable. Enhancement or degradation of QoS is thus applied to specific packets iden-
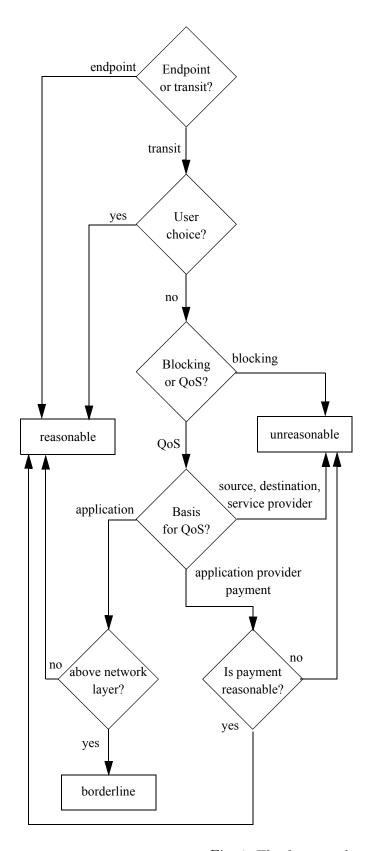
endpoint

Endpoint
or transit?

transit

yes

User
choice?

no

Blocking
or QoS?

blocking

reasonable

unreasonable

QoS

source, destination,
service provider

application

Basis
for QoS?

application provider
payment

no

above network
layer?

Is payment
reasonable?

no

yes

yes

borderline

Fig. 3. The framework

15

tified by the user, for instance if an ISP chose to give enhanced QoS to all packets identified using diffServ codepoints by the user as VoIP.

The last remaining case consists of practices that are applied *at or above the transport layer at transit nodes without user consent and enhance or degrade QoS on the basis of the application.* Practices of this sort use DPI to identify which packets should receive high or low priority or dedicated bandwidth. A common example of this practice is traffic shaping for file-sharing. Because DPI is used (rather than user identification of these packets), this practice violates layering. The question is whether this violation of layering is severe enough to cause this practice to be classified as unreasonable. There are more direct techniques that can be used that rely on user identification of packet priorities and that do not violate layering, as discussed in the next section of this paper. However, because these alternative practices involve different business models that may require some time to be accepted by the public, we recommend classifying any such practice that uses DPI to apply QoS as a borderline traffic management practice that could be used for a limited period of time if properly disclosed in the user contract.

The resulting framework is summarized by the flowchart in Figure 3.

5 Conclusion

We have proposed a framework that can be used to determine whether a traffic management practice is reasonable. The next step is to illustrate how unreasonable practices can be modified to become reasonable and to achieve similar purposes.

References

Comcast Corporation, Feb 2008. Comments of Comcast Corporation before the Federal Communications Commission in the matter of broadband industry practices.

Cox Communications, May 2004. Whitepaper: Voice over Internet protocol: Ready for prime time.

FCC, 2005a. DA 05-543, Madison River Communications Consent Decree. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/DA-05-543A1.pdf.

FCC, 2005b. FCC 05-151, Internet Policy Statement. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-05-151A1.pdf.

FCC, 2007. FCC 07-31, Broadband Market Practices Notice of Inquiry. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-07-31A1.pdf.

FCC, 2008. FCC 08-183, Comcast Order. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08-183A1.pdf.

Free Press, Public Knowledge et. al., November 2007. A petition before the Federal Communications Commission for declaratory ruling that degrading an Internet application violates the FCCs Internet Policy Statement and does not meet an exception for reasonable network management.

Frieden, R., 2006. Network neutrality or bias? – Handicapping the odds for a tiered and branded Internet. In: Proceedings of the 34th Research Conference on Communication, Information and Internet Policy (TPRC). TPRC.

Jordan, S., May 2009. Implications of Internet architecture upon net neutrality. ACM Transactions on Internet Technology 9 (2), 5:1–5:28.

Lehr, W. A., Gillett, S. E., Sirbu, M. A., Peha, J. M., 2007. Scenarios for the network neutrality arms race. International Journal of Communication 1, 607–643.

Packeteer Incorporated, July 2007. Best practices: Monitoring and controlling peer-to-peer (p2p) applications.

Packeteer Incorporated, 2008. Packeteer education customers. Retrived on July 10, 2008.
URL http://www.packeteer.com/customers/education.cfm

Peha, J. M., 2007. The benefits and risks of mandating network neutrality, and the quest for a balanced policy. International Journal of Communication 1, 644–668.

Saltzer, J. H., Reed, D. P., Clark, D. D., 1984. End-to-end arguments in system design. ACM Trans. Comput. Syst. 2 (4), 277–288.

Sandvine Incorporated, December 2004. Session management: BitTorrent protocol, managing the impact on subscriber experience.

Sandvine Incorporated, July 2007. Sandvine DPI-based policy solutions.

Vuze Inc., November 2007. A petition before the Federal Communications Commission to establish rules governing network management practices by broadband network operators.

Weitzner, D., May-June 2008. Net neutrality... seriously this time. Internet Computing, IEEE 12 (3), 86–89.

Title:

A Layered United States Universal Service Fund for an Everything-over-IP World

Author:

Jordan, Scott, University of California, Irvine

Abstract:

Support through the United States Federal Universal Service Fund for high-cost areas has been principally defined in terms of telephone service. Fund growth due to increases in wireless lines and implicit support for broadband infrastructure has created an untenable situation, and fundamental reform is expected. The cause underlying this growth is convergence between the telephone network, wireless networks, the Internet, and cable networks. This convergence will pose additional serious long term challenges to the Fund. This paper proposes a restructuring of the high-cost funds based on a layered model. Both contributions and distributions are focused on network infrastructure, without distinction between voice and broadband. The proposal uses a new definition of communication services to guarantee technology neutrality, and includes service area reform and cost e±ciency measures. This layered approach repositions the Fund for future converged networks.

A Layered United States Universal Service Fund for an Everything-over-IP World

Abstract

Support through the United States Federal Universal Service Fund for high-cost areas has been principally defined in terms of telephone service. Fund growth due to increases in wireless lines and implicit support for broadband infrastructure has created an untenable situation, and fundamental reform is expected. The cause underlying this growth is convergence between the telephone network, wireless networks, the Internet, and cable networks. This convergence will pose additional serious long term challenges to the Fund. This paper proposes a restructuring of the high-cost funds based on a layered model. Both contributions and distributions are focused on network infrastructure, without distinction between voice and broadband. The proposal uses a new definition of communication services to guarantee technology neutrality, and includes service area reform and cost efficiency measures. This layered approach repositions the Fund for future converged networks.

Key words: policy formulation, universal service, Internet

1. Introduction

The United States Universal Service Fund (USF), as established in the 1996 Communications Act, was created to provide access to advanced telecommunications and information services in all regions of the Nation. Universal service, however, has to date been defined principally in terms of telephone services. Correspondingly, the high-cost funds have focused on support to Eligible Telecommunication Carriers (ETCs) to support telephone network infrastructure, and assessments have been placed on interstate revenue from telephone calls.

Communication and computer networks, however, are continuing a long term trend toward integration of physical facilities and the applications they support. Whereas four separate networks historically supported telephone calls, cable video broadcasting, Internet, and wireless communications, increasingly these networks are integrating into a merged infrastructure that supports a wide variety of voice, video, and data applications.

This convergence of physical networks and applications poses several long-term challenges to the Universal Service Fund. As technology has changed, revenue from telephone calls has decreased whereas revenue from video services, Internet access, and wireless services has increased. Furthermore, it is increasingly difficult to distinguish between revenue from intrastate and interstate telecommunication services. USF support to wireless ETCs has increased rapidly, and it is widely believed that this is the principal cause for the rapid growth of the Fund. Differences in the service areas of Incumbent Local Exchange Carriers (ILECs) and wireless ETCs has required new rules to combat cream-skimming. The development of Voice over IP (VoIP) presents a classification challenge, in that debate has started about whether VoIP infrastructure should be supported by the Fund and whether revenue from VoIP services should be assessed for the Fund. As video over IP starts to come into the consumer marketplace, this will pose a similar challenge to classification. In

addition, many policymakers wish to see the USF support broadband access, but there is some reticence to assess broadband access, and many worry that support for broadband will cause a further massive increase in the size of the Fund.

Because of all these problems caused by technological convergence, the fundamental structure of the high-cost funds must be reconsidered. This paper proposes an approach to reformulation of the high-cost funds. In recognition of convergence, a new term *communication services* is introduced to replace *telecommunication services*, based on a layered model, to restrict both contributions and distributions to network infrastructure not applications. Communications services infrastructure encompass the lower layers of all public interconnected networks, whether to support voice, video, or data. Support is thus focused directly on communications network infrastructure, without reference to a standardized basket of telecommunications services that qualify as universal service. Similarly, contributions are required from all communications services, whether to support voice, video, or data.

This technology-neutral approach repositions the USF for future converged networks. In particular, such a layered approach erases the distinction between support for telephone service and broadband access. To resolve the disappearing distinction between intrastate and interstate revenue, the proposal allows the Fund to assess both. To resolve incompatibilities between the service areas of multiple types of network providers, service area reform is implemented using disaggregation. Finally, data is presented that indicates that a significant component of USF growth has been caused by implicit support for wireline broadband; the proposals suggest that allowing policymakers to determine an overall limit on the size of the Fund is more viable in the long-term than artificial distinctions between voice and broadband distributions.

Although many components of the proposal strictly concern elements of the United States USF (e.g. intrastate versus interstate, and USF growth), a layered approach to USF may apply to other countries universal service programs.

Section 2 briefly reviews the related literature, the current debate, and recent Congressional bills. Section 3 briefly states updated goals for universal service. Section 4 proposes a layered approach to high-cost fund distributions, and section 5 proposes cost efficiency mechanisms based on this approach. Finally, section 6, applies this layered approach to Fund contributions.

2. Related Literature

There is a voluminous literature on problems of the Universal Service Fund and proposed modifications. Abernathy (2005) gives a good overview of issues with the Fund posed by the transition from circuit-switch voice services to packet-switched communication services. On the contribution side, she discusses decreasing long distance telephone call revenues, classification of VoIP, and whether the Federal Communications Commission (FCC) has authority to assess the telecommunications portion of information services such as broadband access. On the distribution side, she discusses support for broadband, whether the FCC has authority to support an information service under USF, and the identical support rule. Marcus (2005) discusses the burdens put on the Fund from wireless carriers, and motivations to expand the Fund to VoIP and broadband. Taylor (2007) discusses a future of everything-over-IP communication services in which it no longer makes sense to distinguish between interstate and intrastate, between wireline and mobile, or between the telephone network and the Internet. On the contribution side, he discusses whether contributions should be required from all those who benefit from the network, from infrastructure providers all the way up to content providers. On the distribution side, he argues that under everything-over-IP, there will be no minimum standard universal service package.

There is a large academic literature on the economic efficiency (or lack thereof) of the Fund. An excellent starting point for digesting this literature is Turner (2006). His economic analysis shows that USF assessments produce a financial burden that falls heaviest on rural consumers who make large amounts of long distance calls, due to the elasticity of demand for rural long distance. In contrast, he presents an analysis that concludes that subsidizing broadband would likely produce benefits that exceed costs. He argues for assessments based on numbers and capacity, a broadband mandate on USF recipients, and the use of actual forward-looking costs.

There are also many papers that compare the approaches to universal service taken by the United States and other countries. Schejter (2007) finds that European countries are deploying broadband more quickly than the United States using a system that is more focussed, simple, efficient, and responsive to technology change. Jayakar and Sawhney (2007) examine new universal service policies in countries that have high broadband growth rates and conclude that they offer a multiplicity of services instead of focussing on voice as a universal service, with consumers able to choose the services that they value the most.

A number of papers discuss whether there is a digital divide, and if so the nature of it. Downes and Greenstein (2006) studied availability of dial-up Internet Service Providers (ISPs). They found that in rural counties, population was the single most important determinant of ISP entry, and that entry was largely provided by ISPs with local focus. This leaves open the question of whether broadband Internet will become less universally available than dial-up service due to differences in explicit USF support for local phone versus broadband. Prieger (2003) presents an analysis that quantifies that differential in broadband availability between urban and rural areas, and finds that concentration of rural households is a more important factor than income.

A lively debate has been occurring about whether the USF should support broadband. Weinberg (2001) argues that the distinction between telecommunications services and information services is not appropriate for the USF. After discussing several classification problems with evolving IP-based services, he proposes that USF contributions and distributions should be associated with physical facilities rather than with services. Sicker (2003) applies a layered policy model to regulation of VoIP. With respect to the USF, he argues that only infrastructure should be assessed and subsidized, not application and content providers. The dividing line between infrastructure and applications is intended to be roughly that between distance sensitive and non-distance sensitive components. He also argues that VoIP is an application, not infrastructure. Oh (2006) argues that the FCC's classification of Internet access as an information service hinders USF support for broadband, and that USF distribution should be expanded to non-telecommunication carriers such as municipal Wi-Fi.

There have been several recent attempts in the United States Congress at reformulating the USF. The telecommunications industry has been actively lobbying on the issue. With regard to distributions, Regional Bell Operating Companies (RBOCs) generally receive only limited funding from the non-rural portions of the Fund. Consequently, since they receive less from the Fund than their subscribers contribute, they are lobbying for a *cap* to be placed on the size of the Fund and for cost-efficiency mechanisms such as *reverse auctions*. In contrast, most rural ILECs, who receive more from the Fund than their subscribers contribute, strongly oppose caps or reverse auctions. Increasingly, wireless carriers receive funds from the rural portions of the Fund, but since their subscribers contributions still dominate the distributions to these carriers, they generally support caps and auctions, provided that the caps do not discriminate against wireless over wireline. Currently, both ILECs and wireless ETCs receive support based on the costs of the ILEC (the *identical support* rule); ILECs are lobbying for wireless ETCs to receive support based on their *actual costs*, but wireless carriers generally oppose this approach unless spectrum costs are included in the cost calculation. The cost calculation for rural ILECs is based on *embedded costs*, whereas the cost calculation for non-rural ILECs is based on *forward-looking costs*; the RBOCs and wireless carriers generally want all calculations to be based on the lower forward-looking costs so that the Fund size decreases, but rural ILECs vigorously oppose such a change. Currently, if a rural household subscribes to multiple voice lines (e.g. voice and fax, or landline and wireless), then all of the household's carriers might receive USF distributions. RBOCs want support to be restricted to a single *primary line* in order to decrease the Fund size, but rural ILECs and wireless ETCs oppose this. Currently, rural ILECs can determine the size of the geographic regions within their service area on which support calculations are based. They would like to continue to have this control, but wireless ETCs would like further *disaggregation* so that they can compete in only a portion of these service areas. Rural ILECs would like the USF to explicitly support broadband, but RBOCs oppose explicit support if it will increase the size of the Fund. Opinions are widely varied about requiring that ETCs offer broadband service (called a *broadband mandate*), depending on the form of the mandate. With regard to contributions, there is general agreement that the Federal USF should assess intrastate telecommunications, but disagreement over assessing broadband and VoIP.

The FCC recently placed an interim cap on high-cost fund distributions to competitive ETCs (CETCs),

to control growth of the Fund until more fundamental reforms can be implemented (FCC, 2008a). As of the date of writing, the FCC is currently considering a set of recommendations by the Federal-State Joint Board on Universal Service (Federal-State Joint Board on Universal Service, 2007). The Joint Board recommends that such reforms include conversion of the high-cost funds to three separate funds for ILECs, wireless ETCs, and broadband ETCs, with an overall cap on these funds. Emphasis in the wireless and broadband funds would be placed on deployment of infrastructure in unserved areas. The Joint Board also recommends that only a single provider in each category be supported, potentially using reverse auctions, and that ETCs receive support based on their actual costs rather than on the costs of the ILEC.

The *Universal Service for Americans Act* (U.S. Congress, 2007b) [1] , introduced by Senators Stevens, Lott, and Hutchinson, was based on the Universal Service title of the 2006 *Advanced Telecommunications and Opportunities Reform Act* (U.S. Congress, 2006), which passed out of the Senate Commerce Committee, but never made it to the Senate floor. With regard to distributions, the bill adds a separate $500M fund for broadband providers in unserved areas and creates a separate fund for support in insular areas (e.g. Alaska and Hawaii). The bill does not place a cap on the Fund or implement reverse auctions, maintains identical support, prohibits restriction of support to a primary line, and is silent on embedded versus forward-looking cost and on disaggregation. With regard to contributions, the bill expands the Federal USF base to include intrastate communications, VoIP, and broadband, and expands the State USF base to include interstate communications and VoIP but not broadband. It allows contributions to be based on revenue, numbers, and/or capacity.

The *Universal Service Reform Act of 2007* (U.S. Congress, 2007a), introduced by Representatives Boucher and Terry, is most comprehensive proposal in the House. With regard to distributions, the bill places a cap on the high-cost funds, adopts actual costs (excluding spectrum), prohibits restriction of support to a primary line, defines a wireless carrier's service area as their basic trading area, disaggregates support in the non-rural fund to wire centers, explicitly funds broadband, and includes a broadband mandate. With regard to contributions, the bill expands the Federal USF base to include intrastate communications, VoIP, and broadband, and similarly expands the State USF base to include interstate communications, VoIP, and broadband. It allows contributions based on revenues and/or numbers.

3. Goals of universal service

This section briefly states updated goals for Universal service. These goals will be implemented through a layered approach to high-cost fund distributions in sections 4 and 5, and to fund contributions in section 6.

The principal stated goal of the USF to date has been to increase the deployment of voice service and to equalize the charges for basic phone service between urban and rural areas. The percentage of households with phones peaked at about 96% in the early part of this decade, and has been decreasing since then, partially due to substitution of cell phones for wireline phones (FCC, 2007). Telephone penetration remains significantly lower for households with low income, but USF LifeLine and LinkUp are not addressed in this paper.

Much of the current debate over the USF has shifted from voice service to broadband Internet, since there remain substantial differences between urban and rural deployment of broadband. Rural areas can be more costly to serve, and deployment of new communications technology such as broadband is less developed (U.S. General Accounting Office, 2002; Kruger and Gilroy, 2007). It is widely believed that ubiquitous and rapid adoption of broadband would result in a large increase in jobs and gross domestic product (Kruger and Gilroy, 2007).

The size of the high-cost funds increased from $2.93B in 2002 to $4.29B in 2007, while assessed revenues decreased from $302B to $300B in a similar time period [2] . Due to both the growth of the size of the Fund and the increase on the contribution factor (the USF assessment expressed as a percentage of interstate and

---

[1] All bills from the 110th Congress expired at the end of 2008. However, similar bills are likely to be reintroduced in the 111th Congress in 2009.
[2] 2001 to 2006.

international telecommunications revenue), Congress views the current system as not sustainable (Goldfarb, 2006).

The principal purpose in this proposal is to reposition the USF for an everything-over-IP world. Much of the current debate over the USF concerns whether it should explicitly adopt a goal to increase the deployment of broadband. Based on technological convergence, this proposal accepts this goal and aims to refocus the Fund on a broader set of objectives that includes both voice and broadband. Not only should rural broadband be a goal of the Fund, but the Fund should encourage future communication services.

The second purpose in this proposal is to encourage competition. In recognition that technological convergence is tearing down technology barriers between various providers of communication services, competition should be encouraged for service in rural areas.

The third purpose in this proposal is to tightly focus the program. The current program expresses objectives in terms of access and pricing of telecommunication services to *consumers* but provides subsidies for infrastructure to *carriers*. This conflation of application and infrastructure creates confusion over the purpose and operation of the Fund in reference to new technologies such as VoIP.

The final purpose in this proposal is to make the Fund viable for a period of many years, so that statutes do not have to be rewritten when the next change in technology might make the term *broadband* dated and of limited use.

Throughout this paper, proposed statute language is included that could be included in a Congressional bill. Proposed statute language is formatted indented and in *italics*. These updated goals for Universal Service could be summarized by Findings and Purposes, as are commonly included in bills:

> *SEC. 1. FINDINGS.*
>
> *The Congress finds the following:*
>
> *(1) The Universal Service Fund has contributed to the nearly universal deployment of voice service and the equalization of charges for basic phone service between urban and rural areas.*
>
> *(2) New communication services such as broadband Internet have become a significant enabler of economic growth, and rural areas lag in deployment of such services.*
>
> *(3) The current State and Federal mechanisms used to collect and distribute universal service support are not sustainable in a competitive and rapidly changing technological environment.*
>
> *SEC. 2. PURPOSES.*
>
> *The purposes of this Act are-*
>
> *(1) to broaden the goals of universal service and to encourage the deployment of new and future communications services in rural areas;*
>
> *(2) to encourage competition for communications services, so that rural Americans can have access to modern communications technologies at equitable prices;*
>
> *(3) to ensure the universal service high-cost program is tightly focused on these goals and results in efficient investments; and*
>
> *(4) to reform the universal service contribution base to ensure its long-term viability, to make it sustainable under future changes in technology, and to make it consistent with revised program goals.*

4. Eligible Telecommunications Carriers and Supported infrastructure

This section considers which network infrastructure should be supported by the high-cost funds. The first two subsections discuss current ETC obligations and supported infrastructure and briefly restate the problems with the current approach. The following three subsections propose a layered approach to defining supported infrastructure and service areas. With the new layered model in hand, section 5 considers fund growth. Sections 4 and 5 therefore together address the distribution side of the Fund. Section 6 addresses the contribution side of the Fund.

4.1. *Current ETC obligations and supported infrastructure*

Common carriers can apply to state commissions or the FCC to be designated as an Eligible Telecommunications Carrier (ETC), and thus be eligible to receive universal service funds. In order to be designated as an ETC, a common carrier must offer the services supported by the federal universal service mechanisms, upon reasonable request, throughout the designated service area, either by using its own facilities or by using a combination of its own facilities and resale of another carriers services, and must advertise these services throughout the designated service area. The default service area for an ETC is the study area of the ILEC that serves that region. However, an ETC can be certified for only a portion of an ILEC's rural study area if the ETC is not judged to be disproportionably serving the high-density low-cost subsets of the study area. The FCC interprets reasonable requests as those within the ETC's licensed service area that can be accommodated at reasonable cost.

In addition, an ETC must: (1) provide a five-year plan demonstrating how high-cost universal service support will be used to improve its coverage, service quality or capacity in every wire center for which it seeks designation and expects to receive universal service support; (2) demonstrate its ability to remain functional in emergency situations; (3) demonstrate that it will satisfy consumer protection and service quality standards; (4) offer local usage plans comparable to those offered by the incumbent local exchange carrier in the areas for which it seeks designation; and (5) acknowledge that it may be required to provide equal access if all other ETCs in the designated service area relinquish their designations pursuant to section 214(e)(4) of the Communications Act (FCC, 2005b).

The high-cost funds of the USF provide support for ETCs operating in rural, high-cost, and insular areas. Support is provided through eight funds[3] that differ by recipients (rural ILECs, non-rural ILECs,

---

[3] The eight high-cost funds are:

HCLS: The High-Cost Loop Support fund is available to rural ILECs and ETCs to help with non-traffic sensitive intrastate costs. It reimburses 65%-75% of the ILEC's non-traffic sensitive local loop embedded costs exceeding 115% of a national average cost benchmark. Prior to 2000, support was also available to non-rural ILECs at lower rates, but this support has been phased out by 2006. There is a cap on the fund that adjusts according to the rate of growth in working loops of rural carriers plus the rate of inflation as measured by the Gross Domestic Product Chained Price Index (GDP-CPI).

SNAS: The Safety Net Additive Support fund is also available to rural ILECs and ETCs to help with non-traffic sensitive intrastate costs. It effectively removes the cap on the HCLS fund for those carriers whose costs per loop increase by more than 14% annually. The SNAS fund is capped.

SVS: The Safety Valve Support fund is also available to rural ILECs and ETCs to help with non-traffic sensitive intrastate costs. It effectively removes the cap on the HCLS fund for those carriers that acquire high cost exchanges. The SVS fund is capped.

HCMS: The High-Cost Model Support fund is available to non-rural ILECs and ETCs to help with non-traffic sensitive intrastate costs. In contrast to the HLCS fund which is based on embedded costs of each carrier, the HCMS fund is based on aggregated statewide forward-looking costs. It provides aggregate support to non-rural carriers in a particular state if the total statewide forward-looking costs of intrastate supported services is above a national cost average benchmark. In states that receive HCMS support, funds are allocated to wire centers with forward-looking costs above the benchmark. The amount that a non-rural carrier receives through HCMS depends on the wire center cost and on the number of lines served by the carrier. The fund is not capped.

ICLS: The Interstate Common Line Support fund is available to rate of return regulated carriers (and corresponding ETCs) to help with non-traffic sensitive *interstate* costs. It is gradually replacing Common Carrier Line (CCL) charges, and the amount provided is based on rate-of-return regulation. The fund is not capped.

LTS: The Long-Term Support fund was also available to rate of return regulated carriers that participated in the National Exchange Carrier Association pool (and corresponding ETCs) to help with non-traffic sensitive interstate costs. It reduced CCL charges, and the amount provided was based on rate-of-return regulation. The fund was merged into the ICLS fund in 2004.

IAS: The Interstate Access Support fund is available to price cap regulated carriers (and corresponding ETCs) to help with non-traffic sensitive *interstate* costs. It is gradually replacing CCL charges. Funds are targeted to low density areas, and the amount provided is based on the amount of CCL charges replaced above the Subscriber Line Charge (SLC). The fund has a $650M target size.

LSS: The Local Switching Support fund is available to ILECs with study areas of 50,000 or fewer access lines (and corresponding ETCs) to help with traffic-sensitive switching costs. It reduces traffic-sensitive access charges to inter-exchange carriers, and the amount is based on rate-of-return regulation. The fund is not capped.

ETCs, rate of return carriers, price cap carriers), by geography (intrastate, interstate), and by the type of infrastructure (non-traffic sensitive, traffic-sensitive) (FCC, 2007).

4.2. *Problems*

There are several problems that have been discussed in the literature concerning obligations of an Eligible Telecommunications Carrier and supported infrastructure. The high-cost funds were originally intended to expand the number of residences with wireline phones, and to equalize the rates charged for phone service in urban and rural areas. As a consequence, the obligations of an ETC are principally concerned with phone service, and the distributions of the high-cost funds only explicitly address telephone network costs.

However, since the last revision of USF statutes, the number of wireline phone lines has peaked and the number of wireless phone lines has skyrocketed. The number of wireless lines now exceeds the number of wireline lines, and in many rural areas wireless service has become more cost-effective than wireline service. As a result, many wireless providers have become CETCs and high-cost fund distributions to CETCs have risen from $46M in 2002 to $1137M in 2007 (FCC, 2007). Many blame the growth of the USF almost entirely on this growth in CETCs, although the analysis below indicates blame should be shared with increases in ILEC line costs.

In funds that support rural carriers, distributions are based on the costs of the ILEC (the original provider of wireline phone service) in each ILEC study area. Wireless providers, however, have different service areas (based on spectrum licenses) that do not entirely coincide with ILEC study areas. As a consequence, wireless ETCs have often requested to serve only a portion of an ILEC study area, and have often been accused of cream-skimming (serving only those portions of an ILEC study area that are relatively low cost). In addition, VoIP has become an alternative voice service. Most cable companies now offer phone service using VoIP, and a few have become ETCs to support this service. Similar problems can occur with cable company service areas.

Finally, although the high-cost funds currently *implicitly* support broadband infrastructure, there are active debates over whether USF should *explicitly* support broadband.

The historic distinctions between the telephone network, cable video networks, the Internet, and wireless networks are disintegrating. Integration is creating networks that support a wide variety of voice, video, and data applications. In an everything-over-IP network, many existing distinctions no longer apply. Voice becomes only one of a large number of services that carriers offer. Definition of a standardized basket of universal services thus becomes difficult, since users will differ greatly in which services are of interest to them. Requiring a carrier to offer all services will likely be overly restrictive and inefficient. References to ILEC costs and service areas become arbitrary. Costs become much less distance-sensitive, and thus are less often identifiable as intrastate or interstate. The future will bring challenges to the USF far in excess of today's debate around supporting broadband. In an everything-over-IP network, there will be no distinction between telephone network infrastructure and Internet infrastructure, and hence no clear distinction between costs related to voice and broadband.

The current USF program expresses objectives in terms of access and pricing of telecommunication services to consumers but provides subsidies for infrastructure to carriers. This conflation of application and infrastructure creates confusion over the purpose and operation of the Fund in reference to new technologies.

As an example, consider VoIP service. Should VoIP providers receive USF distributions? Consider the following scenarios:

(a) A rural local exchange carrier offers VoIP over twisted-pair lines, and routes the traffic onto the Public Switched Telephone Network (PSTN).

(b) A rural cable operator offers VoIP over cable, and routes the traffic onto the PSTN.

(c) A rural cable operator offers VoIP over cable, and routes the traffic over the Internet.

(d) A rural local exchange carrier also operating as a cable operator offers VoIP over cable.

———

(e) A company offers VoIP service for a fee to rural consumers who already purchase broadband service from another provider.

(f) A company offers VoIP service to rural consumers who already purchase broadband service from another provider, but charges only for calls routed onto the PSTN.

Which of these VoIP providers should receive USF distributions? Does it matter if the VoIP service provider is a facilities-based carrier or not? If so, what type of facilities count – local loops? lines? routers? Does it matter if the VoIP traffic is routed from the Internet to the PSTN (or vice versa) or if it runs entirely over the Internet? Does it matter if the carrier is a local exchange carrier or a cable operator?

Similar problems arise with broadband service. Which of the following should be eligible for USF distributions?

(a) a rural local exchange carrier offering broadband over Digital Subscriber Line (DSL).

(b) a rural cable operator offering broadband over cable.

(c) a wireless carrier offering fixed broadband in a rural location.

(d) a wireless carrier offering mobile broadband including rural areas.

(e) an Internet Service Provider who offers Internet services in conjunction with a local exchange carrier offering DSL service?

Convergence will only make the issue worse. How should USF address rural local exchange carriers or cable operators who provide video over the same IP networks as broadband? Application providers who offer services that integrate voice, video, and data for specific interactive applications such as gaming? Packages in which the voice component is provided free on top of other higher end services?

The distinction between voice network infrastructure and broadband network infrastructure is quickly disappearing, and USF funds are already being used by many carriers to deploy combined voice and broadband network infrastructure (discussed more below).

4.3. *Layered network architecture*

Much of the confusion regarding what the USF should support emanates from difficulties with classification of evolving communication services. The current program expresses objectives in terms of access and pricing of telecommunication *services* but provides subsidies for *network infrastructure*. It attempts to assess revenue from voice applications in order to subsidize voice infrastructure; in reality, it assesses revenue from voice applications and subsidizes combined voice and broadband infrastructure. Many of the questions above concern whether VoIP or broadband is principally a service or network infrastructure, whether the type of network infrastructure matters, and whether the type of service provider matters.

The principal idea in this paper is to separate this conflation of support for network infrastructure but assessment of revenue from network applications by both support for and assessment of only network infrastructure. Furthermore, the goal is to remove reference to voice (or any specific list of applications) as the principal supported service. This approach relies on the ability to partition the network into applications and infrastructure.

Telephone networks, cable video networks, cellular networks, and the Internet are all based on the concept of a layered architecture. Each network device, and the network as a whole, is abstractly modeled as being composed of a number of vertical layers. Each layer provides certain functionalities.

The reference model for layered architectures is the Open Systems Interconnection (OSI) model, developed by the International Standards Organization. The OSI model is composed of 7 layers, as pictured in figure 1. OSI layer 1, called the physical layer, implements encodes a bit into a physical signal and vice versa. OSI layer 2, the data link layer, translates a packet into a set of bits and vice versa, and implements a set of rules (called a protocol) about which device can transmit when. OSI layer 3, the network layer, is concerned with routing a packet from one network device to the next. OSI layer 4, the transport layer, is concerned with functionality required to form a complete connection between a source and destination, including dealing with lost packets and responding to congestion. OSI layer 5, the session layer, manages an entire communication session, e.g., logging onto a service. OSI layer 6, the presentation layer, concerns data presentation, e.g., file or video compression. Finally, OSI layer 7, the application layer, deals with user

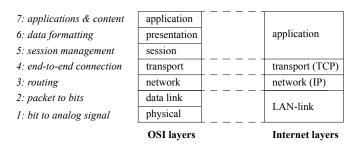|			
7: applications & content	application		
6: data formatting	presentation		application
5: session management	session		
4: end-to-end connection	transport		transport (TCP)
3: routing	network		network (IP)
2: packet to bits	data link		
1: bit to analog signal	physical		LAN-link
	OSI layers		**Internet layers**

Fig. 1. OSI and Internet layered models

applications and other high-level functionality, e.g., web browsing, e-mail, file transfer, file sharing, instant messaging, gaming, etc.

Not every network device contains all 7 layers. Personal computers do contain all 7 layers - the network interface card (e.g., Ethernet card) implements OSI layers 1 and 2, the operating system (e.g., Windows) implements OSI layer 3 and part of layers 4 through 7, and user-installed software implements the remainder of layers 4 through 7. A network router, however, often contains only layers 1 through 3.

Layers 1 through 3 can be thought of as network infrastructure, while layers 4 through 7 can be thought of as network applications. This distinction will be discussed in more detail below as the impact of layered architectures upon communications policy is considered.

Although the OSI model serves as a reference for all network architectures, different networks have modified the model for their own use. As an example, the Internet uses a model with a reduced number of layers, as pictured in figure 1. OSI layers 1 and 2 are combined into a single Internet LAN-link layer. OSI layer 3 is also called the Internet network layer; it includes the Internet Protocol (IP). OSI layer 4 is also called the Internet transport layer; it includes the Transmission Control Protocol (TCP). OSI layers 5 through 7 are merged into a single Internet applications layer.

4.4. *A layered regulation model*

The rationale of this proposal is that the USF should both support and assess physical network infrastructure not applications. More specifically, the Fund should support network infrastructure whose cost per subscriber is sensitive to geographical subscriber density. The question is how to determine which portions of the network are infrastructure and which are applications.

The FCC has provided little useful guidance on this matter. The result of the Computer I (FCC, 1971) and Computer II (FCC, 1980) inquiries was roughly to classify the Internet LAN-link layer as basic or telecommunication services and to classify the IP and higher Internet layers as enhanced or information services. More recently, the FCC classified broadband offered over DSL and broadband over cable networks as information services (FCC, 2002, 2005a). As of the date of writing, the FCC is currently considering adding broadband as USF supported network infrastructure (FCC, 2008a), but not as an assessed application. Although VoIP has not been classified, the FCC is treating facilities-based VoIP providers as infrastructure but non facilities-based VoIP as an application. The FCC has yet to classify other IP-enabled services.

The goal is to specify which network infrastructure to support. Supported network infrastructure should consist of components whose cost per subscriber is sensitive to geographical subscriber density. Fortunately, this is easy to identify using the OSI model. OSI layers 1 through 3 must be implemented at every network hop from source to destination; these layers therefore include those network mechanisms that must be provided by each carrier on their portion of the network, namely those that are sensitive to subscriber density. In contrast, OSI layers 4 through 7 need only be implemented at the endpoints (the source and destination); these layers include mechanisms that are not sensitive to subscriber density. USF supported infrastructure should therefore be restricted to OSI layers 1 through 3. As a result, supported infrastructure will include network infrastructure that is high barrier-to-entry, and exclude applications that are low barrier-to-entry.

This layered approach works equally well with both circuit-switched and packet-switched networks, including the PSTN, the Internet, wireless networks, and cable networks, since all are based on layered models.

9

Communication at or below the network layer does not meaningfully change the content of the information, whereas some protocols above the network layer create, store, or change the presentation of information, so this definition roughly complies with the original intent (but not more recent FCC implementation) of the distinction between telecommunication and information services.

Note, however, that the layers of a service are defined by its functionality, not by the device that implements the service. Many implementations of the functionality within a specific layer will combine both hardware and software[4]. Supported infrastructure will thus include all infrastructure required for transport functionality in both telephone and broadband networks (e.g. loops and switches for circuit-switched voice and video conferencing, and loops and switches for packet-switched voice, video, and data), but exclude higher-level functionality (e.g. call-forwarding, voicemail, IP-enabled service, email storage, and other applications).

Services offered to the consumer require functionality at multiple layers. The terms *broadband service* and *Internet access* are often used imprecisely to mean various functionalities at various layers. Only the portion of broadband service at network layer and below should be supported. This includes the basic high barrier-to-entry functionality that allows a subscriber to access the Internet, but excludes value-added low barrier-to-entry services (e.g. email or web-hosting) that are competitively offered by many application providers. Similarly, only the portions of VoIP at network layer and below should be supported. This is consistent with the current implementation, which allows support only for facilities-based VoIP providers.

This use of layers to delineate infrastructure from applications differs from other layered approaches taken in the USF literature. Weinberg (2001) proposes associating both USF high-cost contributions and distributions with physical facilities, defined in a layered manner. He does not, however, specify which layers to define as physical facilities. Sicker (2003) proposes a four layer model consisting of content, applications, transport, and access that does not necessarily map directly into the Open Systems Interconnection (OSI) model. With respect to the USF high-cost funds, he proposes that both the contribution base and the distribution base be restricted to transport and access providers. He classifies VoIP as an application, and thus excludes VoIP from the USF. Although the proposal in this paper agrees with assessing services at the same layers to which distributions flow, the idea is to use the OSI model to delineate the layers, and exclude the OSI transport layer (which resides just above the network layer) from USF since the cost of this layer is not density dependent, whereas Sicker (2003) would probably include the OSI transport layer in his transport layer. This proposal also disagrees with the classification of VoIP solely as an application layer service, and includes within USF those portions of VoIP at the network layer and below.

This delineation of infrastructure from applications is potentially of use more broadly within telecommunications policy. Wherever it has been of value to consider the distinction between telecommunication services from information services, this new definition of communication services may help make a case for substantially different regulation of infrastructure than of applications. In particular, Jordan (2007) uses a similar distinction to argue that net neutrality can be implemented by requiring an open interface between infrastructure and applications. Whereas many papers in the literature have argued for similar horizontal regulation (see e.g. (Lessig, 2001; Werbach, 2002; Solum and Chung, 2003; Whitt, 2004)), most of these layered models combine OSI layers 3 and 4 (the Internet network and transport layers) into a single logical layer, whereas the approach taken here proposes that OSI layer 3 should be classified as infrastructure and OSI layer 4 should be classified as applications.

This layered approach can be formalized using new definitions. A new term *communication services* is introduced to replace *telecommunication services* to restrict both contributions and distributions to network infrastructure not applications:

SEC. 3. DEFINITIONS.

Section 3 of the Communications Act of 1934 (47 U.S.C. 153) is amended-

(1) by inserting after paragraph (10) the following:

"**(10A)** COMMUNICATIONS INFRASTRUCTURE.- The term 'communications infrastructure' means all network infrastructure required to support communication services.

---

[4] *Firmware* is software that is embedded within a hardware device; while it usually is used to implement functionality at OSI layers 1 through 3, it can be used to implement higher layer functionality at an endpoint. *Middleware* is software that allows multiple processes to interact across a network; it is implemented in the Internet application layer.

"(10B) COMMUNICATIONS SERVICES.- *The term 'communications services' means all services-*
"(A) *over a network that uses a public right-of-way;*
"(B) *over a portion of the network that interconnects to other networks; and*
"(C) *that reside at or below the network layer or are required to manage the network."; and*
(2) *by inserting after paragraph (29) the following:*
"(29A) NETWORK LAYER.- *The term 'network layer' means the third layer of the 7-layer Open Systems Interconnection Model responsible for message addressing and for routing information within the network, including routing within the telephone network and including the Internet Protocol within the Internet.".*

Start with the term *communications services* [5]. Item (C), together with the definition of *Network Layer* implements the delineation between infrastructure and applications discussed above. Item (A) restricts USF support to networks that use a public right-of-way, and is assumed to include the rights-of-way used by ILECs, cable companies, and spectrum used by wireless providers [6].

Item (B) ensures that USF supports infrastructure only when it is available to support applications that can be provided by any party, not just by the communications service provider that provides subscribers end access to the network. It can be integrated with network neutrality or other open access requirements to strengthen this intention (see e.g. (Jordan, 2007).) As a result, companies that offer services entirely over their own networks are excluded from the definition of communication services, and therefore from participation in USF; the most common example of this would be cable companies that offer services entirely over their own networks through title VI.

The term *communications infrastructure* is simply defined as all network infrastructure required to support communications services.

In the following sections, these terms will be used to limit the scope of both USF distributions and contributions. Support is thus focused directly on communications network infrastructure, without reference to a standardized basket of telecommunications services that qualify as universal service. Similarly, contributions are required from all communications services, whether to support voice, video, or data. This technology-neutral approach repositions USF for future converged networks. In particular, such a layered approach erases the distinction between support for telephone service and broadband access.

4.5. *Proposal for supported infrastructure*

As discussed above, it has been less than clear which VoIP components should be eligible for USF distributions. In addition, if broadband is to be supported, it is also less than clear which broadband components should be eligible for support. Future IP services that integrate telecommunication components will make the situation even murkier. The distinction between voice network infrastructure and broadband network infrastructure is quickly disappearing. The evolving network infrastructure will support integrated voice, video, and data applications.

This evolution toward an everything-over-IP network makes the current definition of universal services problematic. Under section 254(c) of the Communications Act, universal service is "an evolving level of telecommunications services that the Commission shall establish periodically under this section, taking into account advances in telecommunications and information technologies and services." The definition of supported services must take into consideration whether a service has been subscribed to by a substantial

---

[5] One note of caution: the term *telecommunications service* pervades the Communications Act of 1934. Corresponding changes from *telecommunications service* to *communications service* could be carefully considered elsewhere, but are not required for this proposal.

[6] This is a bit tricky, as different parts of the Communications Act apply to different types of communication. In particular, since *telecommunication service* is defined as "the offering of telecommunications for a fee directly to the public, or to such classes of users as to be effectively available directly to the public, regardless of the facilities used", it may be of value to modify the definition of communication services proposed here to consider public services rather than just use of public right-of-way. However, this may cause two problems. First, it is preferable to have a definition that does not rely on the end service sold to the user, since many communication services will not be directly sold to the public but instead support applications that are sold to the public. Secondly, networks that do not use a public right-of-way may be regulated in a very different manner.

majority of residential customers; this provision has delayed inclusion of broadband into the definition. However, as voice, data, and video services merge, how will the FCC determine which should be included in the definition of universal service. Should adoption by a minority of residential customers disqualify a service? What will they do about network infrastructure shared between a wide variety of services, some of which qualify and some of which do not?

The only viable long-term strategy for the USF is to support the integrated network infrastructure that underlies all combinations of such applications. Convergence should be embraced and the USF should support all communication services. These services may be offered on networks engineered to provide superior service for a targeted set of applications, or more likely be offered on broadband IP networks that support a wide variety of integrated applications. This approach removes the need for a definition of *Universal Service*, and hence section 254(c) of the Communications Act could probably be removed. At a minimum, 254(c)(1)(B) should be deleted to remove the requirement that universal services be subscribed to by a substantial majority of residential customers.

The groundwork for this approach has been laid with the introduction of the terms *communication services* and *communication infrastructure* in the previous section. Using this layered approach, the USF can directly support communications infrastructure, without reference to a standardized basket of telecommunications services that qualify as universal service:

SEC. 4. UNIVERSAL SERVICE HIGH-COST FUNDS DISTRIBUTION.

(a) QUALIFYING UNIVERSAL SERVICE INFRASTRUCTURE LIMITATIONS.- In carrying out section 254(b)(3) of the Communications Act of 1934 (47 U.S.C. 254(b)(3)), the Federal Communications Commission, acting through the Universal Service Administrative Company, shall ensure that-

(1) support under the Universal Service Fund high-cost funds is made available exclusively for communications infrastructure in rural, insular, and high-cost areas;

(2) in areas in which only a portion of the network infrastructure is communications infrastructure, support under the high-cost funds is prorated to reflect the ratio of communications infrastructure to all network infrastructure;

This proposed statute language requires that the USF high-cost funds be used exclusively for *communications infrastructure* in high-cost, rural, and insular areas. As discussed below, USF funds are already being used by many carriers to deploy combined voice and broadband network infrastructure. Support for broadband is therefore currently implicit. Since communications infrastructure can be used to support both voice and broadband service, section 4(a)(1) makes support for broadband infrastructure explicit, but it does this in a technology-neutral manner that will survive future changes in technology. Such an approach would allow for a reduction in the number of high-cost funds. It would also elegantly remove the distinction between wireline voice, wireless voice, and broadband, and therefore remove the need for the three separate wireline, wireless, and broadband funds proposed by the Joint Board.

There are likely to be situations when network infrastructure is shared between communications services and non-communications services, e.g. a local access network that is used to support voice, broadband, and video. If the video is offered over a portion of the network that resides entirely in the access providers domain, as is likely if it is offered under title VI, then it does not qualify as a communications service. Correspondingly, under 4(a)(2), the carrier is eligible for funds only for the portion of the infrastructure that qualifies as communications infrastructure.

A summary of this proposal, in comparison to current law and Congressional bills, is presented in figure 2. None of the congressional bills explicitly goes this far toward integrating voice and broadband. The Stevens bill (U.S. Congress, 2007b) creates a separate $500M fund for broadband in unserved areas, with a single broadband provider in each area qualifying for these funds. It also creates a separate fund for insular areas, but does not otherwise modify USF distributions. The Boucher/Terry bill (U.S. Congress, 2007a) expands the definition of universal services to include broadband, and therefore explicitly includes support for broadband in USF distributions. It also includes a mandate that ETCs offer broadband (at a rate of at least 1Mbps within 5 years). None of these bills uses a layered approach.

In contrast, this proposed approach is technology-neutral. It erases the distinction between support for telephone service and broadband access, and it should survive future changes in network technology and services.

Issue	Subissue	Current law	Stevens	Boucher/Terry	Joint Board proposal	This proposal
Contributions	assessed services	interstate & international telecommunication services	adds intrastate, VoIP, broadband	adds intrastate, VoIP, broadband	silent	adds intrastate, communication services, including facilities-based VoIP and broadband
	methodology	telecommunication services revenues	revenue, numbers, and/or capacity	revenues and/or numbers	silent	revenues associated with communication services infrastructure
	state USF	intrastate telecommunication services	adds interstate, VoIP	adds interstate, VoIP, broadband	silent	adds interstate; assess all communication services based on primary place of use, including infrastructure portions of facilities-based VoIP and broadband
High Cost Distributions	ETC obligations	must offer universal services (see description in text)	silent	broadband mandate; all ETCs must act as a COLR	obligation associated with fund	replaces mandated services with distributions for services rendered and for costs related to COLR obligations
	supported infrastructure	see list of current funds	as current, plus adds separate fund for insular areas	as current	single wireline, wireless, and broadband provider per area	support all communication services infrastructure, defined by layers
	broadband	implicit support for shared voice/bb infrastructure	supports single facilities-based provider per unserved area through a separate fund	supports	primarily for construction in unserved areas	supports portion included in communication services infrastructure
	primary line	silent	prohibits	prohibits	silent	oppose; support based on customers served & services rendered
	aggregation	rural: several choices by rural ILEC; non-rural: wire center	silent	silent	silent	wire center, with recommendation for smaller granularity
	service area	case by case cream-skimming analysis if serving less than rural study area	silent	wireless service area = basic trading area	silent	subset of wire center in which ETC is licensed
	actual costs	ILEC costs	silent	actual costs (excluding spectrum)	actual costs	actual costs
	cap	on some funds; some caps indexed	silent	cap on total high-cost funds indexed by inflation & line growth	cap on each fund, decreasing over time	recommend cap indexed by communication services revenue
	reverse auctions	not used	silent	silent	recommended	oppose

Fig. 2. Summary of recommendations

13

4.6. *Service Area Reform*

The normal situation is that an eligible telecommunications provider will provide service upon request in the entire study area of a rural local exchange carrier (RLEC). For wireless communication providers, however, this can be a problem since they may have spectrum licenses over only a portion of a study area. It is thus increasingly common for an ETC applicant to request to serve only a subset of a RLEC's study area.

Currently, however, the support level is determined by the RLEC's average per line cost over the study area. Since the per-line cost often varies substantially over a study area, inversely proportional to subscriber density, this creates a potential for *creamskimming*, which occurs when ETCs serve a disproportionate share of the low-cost, high revenue customers in a RLECs study area. When an ETC applies for USF distributions in a subset of a RLEC's study area, the FCC conducts a creamskimming analysis that compares the population density of each wire center in which the ETC applicant seeks designation against that of the wire centers in the study area in which the ETC applicant does not seek designation (FCC, 2005b). Based on this analysis, the Commission denies designation if it concludes that the potential for creamskimming is contrary to the public interest.

The root cause of this issue is again technological convergence. Since different providers have deployed network infrastructure in different geographical regions, reference to average costs over the LEC's service area is now outdated.

The solution requires addressing both the definition of a service area and the geographical units over which costs are calculated. The two are connected – the service area determines the region in which an ETC is required to provide service upon reasonable request, and the cost basis determines the extent to which an ETC has motivation to cover an area.

One approach that would solve both issues is to disaggregate costs to very small geographical units. If wireline, wireless, and cable ETCs service areas could all easily be described in terms of these small geographical units, then reference to any particular type of provider's service area can be avoided. Furthermore, if costs are disaggregated to small enough units so that costs accurately represent residential density, then the requirement to serve an entire service area can also be removed; the financial incentive should be enough to attract providers to each geographical unit. This approach would also easily accommodate service areas that change whenever a provider obtains or loses access to public right-of-way or spectrum in different areas.

However, disaggregation to very small geographical units may increase the complexity of accounting too much to be politically acceptable. A more moderate approach that can still make substantial progress on both problems (service area definition and cost basis) is to recognize that three types of providers are currently recognized in the Communications Act: wireline, wireless, and cable. Reliance on the ILEC's study area can be removed by allowing wireless and cable ETCs to serve a subset of a study area if it serves all areas for which it is licensed:

SEC. 5. UNIVERSAL SERVICE AREA REFORM.

(a) *MODIFICATION OF SERVICE AREA DEFINITION FOR WIRELESS AND CABLE.- Section 214(e)(5) of the Communications Act of 1934 (47 U.S.C. 214(e)(5)) is amended by adding at the end "In the case of a commercial mobile service (as defined in section 332(d)) that is an eligible telecommunications carrier, the term 'service area' may, at the election of such carrier, be all areas within an incumbent local exchange carrier's service area for which the commercial mobile service has spectrum licenses. To the extent that a cable operator (as defined in section 602(5)) is treated as a common carrier and, as such, is an eligible telecommunications carrier the term 'service area' may, at the election of such carrier, be all areas within an incumbent local exchange carrier's service area for which the cable operator has a franchise (as defined in section 602(9)).".*

Subsection 5(a) codifies the service areas for wireless carriers and cable companies offering communication services. Such providers may elect a service area that consists of the subset of an ILEC's service area for which the provider has spectrum or franchise licenses.

This does not completely solve the cream-skimming issue, since the CETC's licensed area may be lower-cost than the ILEC's average cost over its study area. A moderate approach to solving this problem has been

suggested in past years using *wire center averaging*, in which USF distributions are based on the average cost in each wire center. An example of such an approach is taken from the *Rural Universal Service Equity Act of 2003* (U.S. Congress, 2004):

> **(b)** *FCC TO REVISE SUPPORT LEVEL.- Section 254 of the Communications Act of 1934 (47 U.S.C. 254) is amended by adding at the end the following new subsection:*
>
> > *"(m) UNIVERSAL SERVICE SUPPORT FOR HIGH-COST AREAS.-*
> >
> > > *"(1) CALCULATING SUPPORT.- In calculating Federal universal service support for eligible telecommunications carriers that serve rural, insular, and high-cost areas, the Commission shall revise the Commission's support mechanism for high-cost areas to provide support to each wire center in which the carrier's average cost per line for such wire center exceeds the national average cost per line by such amount as the Commission determines appropriate for the purpose of ensuring the equitable distribution of universal service support throughout the United States.*
> > >
> > > *"(2) HOLD HARMLESS SUPPORT.- In implementing this subsection, the Commission shall ensure that no State receives less Federal support calculated under paragraph (1) than the State would have received, up to 10 percent of the total support distributed, under the Commission's support mechanism for high cost areas as in effect on the date of the enactment of this subsection.*
> > >
> > > *"(3) LIMITATION ON TOTAL SUPPORT TO BE PROVIDED.- The total amount of support for all States, as calculated under paragraphs (1) and (2), shall be equivalent to the total support calculated under the Commission's support mechanism for high cost areas as in effect on the date of the enactment of this Act.*
> > >
> > > *"(4) CONSTRUCTION OF LIMITATION.- The limitation in paragraph (3) shall not be construed to preclude fluctuations in support on the basis of changes in the data used to make such calculations.*
> > >
> > > *"(5) IMPLEMENTATION.- Not later than 180 days after the date of the enactment of this Act, the Commission shall complete the actions (including prescribing or amending regulations) necessary to implement the requirements of this subsection.*
> > >
> > > *"(6) COMMISSION'S SUPPORT MECHANISM FOR HIGH-COST AREAS DEFINED.- In this subsection, the term 'Commission's support mechanism for high-cost areas' means the mechanism described in sections 54.309 and 54.311 of the Commission's Regulations (47 C.F.R. 54.309 and 54.311, respectively).".*

Because wire centers are not small enough geographical units to remove the incentive to cream-skim, the requirement in FCC regulations to serve all residential customers within a service area must be maintained. With that provision, the combination of 5(a) and 5(b) ensures that ETCs cannot cream-skim, while removing the burden upon State commissions and the FCC to make case by case determinations (and the resulting inconsistencies).

It should be noted, however, that although this moderate approach puts wireline, wireless, and cable providers on an almost equal footing, it does maintain reference to the topology of the telephone network through the use of the term "wire center". In the future, it may be preferable to move to the use of smaller geographical units to remove this dependence.

The Boucher/Terry bill defines a communication provider's service area as a geographic area that aligns with the area in which the provider is licensed to provide service; for wireless service providers, this is explicitly defined as the company's basic trading area. This approach is similar to section 5(a). The bill does not address disaggregation. The Stevens bill does not include service area reform. These differences are displayed in figure 2.

5. Fund Growth

The previous section proposed a layered approach to defining supported infrastructure and service areas. With the new layered model in hand, this section continues discussion of the distribution side of the USF, but now focusses specifically on Fund growth. Section 6 addresses the contribution side of the Fund.

The first subsection presents a brief analysis of growth of the high-cost fund over the last 5 years. The

second subsection reviews challenges posed by this growth and by technological convergence. The third subsection proposes cost saving measures.

5.1. *Analysis*

USF support to wireless ETCs has increased rapidly, and it is widely believed that this is the principal cause for the rapid growth of the USF. In addition, many believe that USF support for broadband would cause the size of the fund to double or triple. In this subsection, data is presented that indicates that a significant component of USF growth has been caused by implicit support for wireline broadband. On this basis, policymakers' determination of an overall limit on the size of the Fund is more viable in the long-term than artificial distinctions between voice and broadband distributions.

This analysis of USF growth is based on high-cost funds distributions during the fiscal years 2002 through 2007 (FCC, 2007)[7]. During this 5 year period, the high-cost funds experienced real annual growth of approximately 5.3%, for a total growth of \$974M on a 2002 base of \$3317M. Of this \$974M increase, approximately 55% (\$537M) can be attributed to an increase in per line costs above the SLC, mostly due to an increase in the rural National Average Cost per Loop (NACPL) from \$289/loop to \$335/loop. Approximately 51% (\$501M) can be be attributed to an increase in the number of lines supported by the USF. (The remaining -7% is unaccounted for.)

The increase in per line costs above the SLC contributes to approximately 3.0% real annual growth in the high-cost funds. This increase can be attributed to upgrading of local loops to support broadband. Since the HCLS and HCMS funds are indexed by line growth and inflation, most of the increase in the size of the high-cost funds attributed to increases in per line costs appear as growth in the ICLS fund due to rate of return regulation.

The increase in the number of lines supported by the USF contributes to approximately 2.9% real annual growth in the high-cost funds. Almost all of this growth is due to an increase in the number of rural lines supported. During the 5 year period considered, there was an annual decrease of approximately 2.7% in the number of rural lines on wireline networks (from 21.9M to 19.2M), so all of this growth (plus some) is due to increases in supported wireless lines.

5.2. *Problems*

There are several problems that have been discussed in the literature concerning growth of the USF. As discussed above, the high-cost funds were originally intended to expand the number of residences with wireline phones, and to equalize the rates charged for phone service in urban and rural areas. As a consequence, the Fund was originally designed to adjust the contributions to match required distributions. Furthermore, required distributions were calculated on the basis of ILEC wireline costs, which it was presumed, would increase no faster than the growth in lines.

However, the Fund has experienced rapid growth in the past 10 years. The high-cost funds have grown (in nominal dollars) from \$1264M in 1997 to \$2935M in 2002 to \$4290M in 2007. Continued growth at these rates is considered politically unacceptable. As a consequence, the FCC over time has placed caps on many of the funds.

Payments from high-cost funds to CETCs have grown (in nominal dollars) from \$0M in 1997 to \$46M in 2002 to \$1137M in 2007. Many assume that this increase, driven largely by increases in payments to wireless providers, is the principal cause for growth of the overall high-cost funds. As a consequence, the FCC has placed an interim cap on high-cost fund distributions to CETCs (FCC, 2008a).

---

[7] Since distributions in 2002 through 2007 were based on costs in 2000 through 2005, loop costs and number of loops are analyzed for the period from 2000 through 2005. All dollar figures are indexed for inflation during 2000 through 2005 and are given in 2007 dollars. Costs per line are measured by the rural National Average Cost Per Line (rNACPL). The decrease in real IAS distributions to CETCs during the period are attributed to the CAP on IAS. Growth in CETC distributions not associated with increases in cost per line or the IAS cap are attributed to CETC line growth.

As discussed in the previous subsection, however, the growth in high-cost funds has been caused both by increases in the number of lines supported and by increases in the real cost per line. Several proposals are under consideration. One general approach is to place a cap on the overall Fund, or on the high-cost funds that are not currently capped. A few approaches are focussed on reducing the cost per line. One idea is to use reverse auctions, in which carriers would bid on how much support they require in order to provide universal service in a particular geographical region. Another idea is to remove the identical support rule, which in rural study areas awards support on the basis of the embedded costs of the ILEC, and to calculate support on the basis of the actual costs of each provider. Yet another idea is to change the support calculation in rural study areas from embedded costs to forward-looking costs, which are presumed to be generally lower. Other approaches are focussed on reducing the number of lines supported. The most common of these is to restrict support to a single primary line per household.

Finally, the debate over whether the USF should explicitly support broadband is intimately tied to the size of the Fund. Opponents often believe that adding support for broadband will dramatically increase the size of the Fund, and that including a broadband mandate will increase the cost per line.

Additional problems will arise as the technology progresses toward everything-over-IP, rendering distinctions between wireline communications, wireless communications, telephone networks, and the Internet less meaningful. Reference to a particular provider (e.g. ILECs) and to its costs become increasingly arbitrary. So do references to a particular service (e.g. voice or broadband).

As a consequence, some of the ideas proposed are not feasible in the long run. As of the date of writing, the FCC is currently considering the use of reverse auctions (FCC, 2008c). Reverse auctions would most likely require definition of a standardized basket of universal services to be supported. The FCC is tentatively concluding that universal service be initially defined as the current universal service obligations plus broadband at a rate of at least 768kbps in one direction. In addition, the geographical area must also be specified in the auction. The FCC is tentatively concluding that ILEC study areas be used. However, if in the future users will be able to choose among a wide variety of IP-enabled services, definition of a standardized basket of universal services becomes very difficult. Furthermore, different providers have overlapping service areas, which makes design of an auction much more challenging.

Similarly, restriction to a primary line or to a single auction winner per geographical area is awkward. Already, users often choose different providers for voice, data, and video services. As users may choose to receive various services over various interfaces, it is strange to specify a single ETC per region that must provide all services.

5.3. *Proposed Efficiency Measures*

As noted, about approximately half of the real growth of the USF high-cost funds in the last 5 years was caused by increases in real per line costs above the SLC, with the other half caused by an increase in the number of lines served. To address this growth, several widely discussed cost-saving measures can be incorporated into this proposal.

The first measure is to provide high-cost fund support on the basis of subscribers rather than on the number of residences in the geographical region. This approach prevents growth caused by multiple networks serving a service area, except when customers receive service from multiple providers. One potential problem with providing support per-subscriber is that ILECs have carrier-of-last-resort obligations that require investment for customers not served. As of the date of writing, the FCC is currently considering whether there should thus be explicit support for carrier-of-last-resort obligations in addition to per-subscriber support.

The second measure is to switch from providing support to CETCs on the basis of ILEC costs to providing support to each ETC on the basis of its own costs. Such an *actual cost* approach is also tentatively proposed by the FCC (FCC, 2008b). Maintaining a reference to ILEC costs (as is currently done) requires an evolving set of predefined services and presumes that the ILEC is the reference for each such service, which is contrary to the benefit of using *communication services* as a stable technology-neutral definition. Switching to actual costs, therefore, has not only the benefit of the expected cost savings, but also of removing reference to a particular provider.

One additional measure is proposed here that has not been discussed in the literature. Support should be prorated on the basis of services rendered. For instance, if a carrier's communications infrastructure is used to provide both voice and broadband, but the customer only elects to purchase voice service, then the carrier should receive support only a prorated portion of the infrastructure. This prevents growth caused by customers receiving different services from multiple providers, and is required in this proposal since support is provided for communications infrastructure independent of a specified set of universal services.

These three measures would require substantial accounting to determine the support due to each ETC. However, the FCC does not seem to view this as a substantial barrier to the use of actual costs or per-subscriber support (FCC, 2008b,c).

These measures can be implemented using this statute language:

SEC. 4. UNIVERSAL SERVICE HIGH-COST FUNDS DISTRIBUTION.

(a) QUALIFYING UNIVERSAL SERVICE INFRASTRUCTURE LIMITATIONS.- In carrying out section 254(b)(3) of the Communications Act of 1934 (47 U.S.C. 254(b)(3)), the Federal Communications Commission, acting through the Universal Service Administrative Company, shall ensure that-

(3) support under the high-cost funds is provided in a nondiscriminatory manner

(A) on the basis of customers actually served;

(B) prorated on the basis of services received;

(C) as necessary to meet carrier-of-last-resort obligations; and

(D) based on the actual cost of providing, maintaining, and upgrading facilities and services in the service area for which such support is intended.

(b) PHASE-IN OF CHANGES.- In carrying out subsection (a), the Commission shall ensure that the changes required in the Universal Service Fund high-cost funds required by this Act are implemented on an incremental basis over a period of not more than 5 years.

In addition to these measures, more aggressive approaches to control of USF growth could be considered. Fund growth due to increases in per line cost could be eliminated by placing a statutory cap on the per line cost. A regulatory approach to accomplish similar purposes could be implemented by placing inflation plus line growth caps on the ICLS and LSS funds, similar to the current cap on the HCLS fund. In contrast, one could consider a cap on the high-cost funds in terms of a percentage of communications services revenue, to allow for growth in infrastructure to match growth in services. Any of these caps, however, might require eliminating the requirement that rates charged in rural areas be comparable to rates charged in urban areas, ensconced in Sec. 254(b)(3) of the Communications Act.

The Boucher/Terry bill includes a few of these approaches. The bill requires that all high-cost fund support be based on actual costs, exclusive of the cost of acquiring spectrum, or on forward-looking costs. It also requires all ETCs to act as carriers of last resort (whereas this proposal provides support for such obligations but does not interfere with the current designation process). The bill removes the individual caps on the HCLS, SNAS, and SVS funds, and replaces them with a cap on the total of the high-cost funds that adjusts according to the GDP-CPI. The Stevens bill does not include any such provisions. Both bills include prohibitions on restriction of support to a primary line. None of these bills explicitly address costs when infrastructure is shared between multiple services or when customers subscribe to a subset of these services. These differences are displayed in figure 2.

6. Contributions

This section turns to the contribution side of the USF. The first subsection briefly recounts the problems with USF assessments. The second subsection considers how to revise the contribution base to survive technological convergence. The third subsection considers the relationship between contributions to the Federal and State universal service funds.

6.1. *Problems*

Consider the contributions side of the USF. Interstate and international telecommunications revenues, which currently serves as the basis for Federal USF assessments, peaked in 2001 (FCC, 2007). In contrast, revenues from video services (cable tv, satellite tv), Internet access (cable modem, DSL), and wireless communications (cell phone, wireless Internet) have continued to increase substantially. As has commonly been observed, this makes reliance on interstate and international telecommunications revenues unsustainable.

There are two separate issues here. First, the distinction between intrastate and interstate telecommunications revenue is becoming problematic. A few decades ago, the cost of a telephone call had a substantial distance-dependent component. Today, however, most communications services are not priced by distance, but only by time. Divisions of revenue between intrastate and interstate are increasingly arbitrary. A commonly suggested fix to this problem is to add intrastate telecommunications services to the assessment basis, and this solution is proposed below.

A second issue, however, is the distinction between telecommunication services and information services. Currently, every telecommunications carrier that provides interstate telecommunication services is required to contribute to the USF (except for de minimis contributions). The bulk of these contributions come from wireline and wireless telephone calls. Whether VoIP providers and/or broadband providers should contribute to the USF remains under debate. The FCC has tentatively agreed that facilities-based VoIP providers must contribute but that broadband providers do not. However, the future of these decisions is unclear.

In addition, as noted above, there is currently a mismatch between defining the contribution base according to the applications included in universal services and defining the distribution base according to supported infrastructure. This mismatch leads to accusations that the Fund is not neutral, either by favoring some providers over others or by some technologies over others. Such problems will get worse as technological convergence increases the variety of applications supported over a merged telephone network and Internet.

6.2. *Neutrality*

Some have suggested adding VoIP and/or broadband services to the USF assessment basis. However, consider the following services:

(a) A local exchange carrier offers VoIP over twisted-pair lines, and routes the traffic onto the Public Switched Telephone Network (PSTN).
(b) A cable operator offers VoIP over cable, and routes the traffic over the Internet.
(c) A company offers VoIP service for a fee to consumers who already purchase broadband service from another provider.
(d) A company offers integrated IP-based voice and video service for a fee to consumers who already purchase broadband service from another provider.
(e) A company offers gaming service for a fee to consumers who already purchase broadband service from another provider, and includes voice service for free.

Which of these are telecommunications carriers that provide telecommunications services? Which should be classified as VoIP service? Which should be classified as broadband service? Which revenue should be assessed? Adding VoIP and/or broadband to the assessment basis leaves it to the FCC to classify services one by one, and the FCC's actions with regard to service classifications has been less than reassuring. In addition, as a wider range of applications is supported by a common networking infrastructure, it is likely that they will often be sold in packages, and it will be increasing difficult to determine whether the associated revenue falls within the assessment basis.

Some have suggested that all information services should contribute based on their reliance on telecommunication services. This approach would solve the problem of service classification, but it would be both infeasible and politically impossible to assess revenue from *all services* that rely on the Internet.

USF should assess services not based on the type of application (e.g. voice or data), nor on the type of network supporting that application (e.g. PSTN and/or Internet), but instead on the layer of the service. The purpose of the Fund is to support infrastructure in rural, insular, and high cost areas. As discussed in

the previous section, with regard to distributions, this should be interpreted as supporting *communications infrastructure* (as defined above). A symmetric approach, with regard to contributions, would be to require contributions from all *communications services*:

SEC. 6. UNIVERSAL SERVICE CONTRIBUTIONS.

(a) FEDERAL USF CONTRIBUTIONS.- Section 254(d) of the Communications Act of 1934 (47 U.S.C. 254(d)) is amended to read as follows:

"(d) CONTRIBUTION BY COMMUNICATIONS SERVICE PROVIDERS.- Every provider of intrastate, interstate, or international communications shall contribute, on an equitable and nondiscriminatory basis, to the specific, predictable, and sufficient mechanisms established by the Commission to preserve and advance universal service. The Commission may exempt a provider, or class of providers, of communications services from this requirement if the provider's communications services are limited to such an extent that the level of such provider's contribution to the preservation and advancement of universal service would be de minimis.".

Subsection 6(a) changes the USF contribution base (i) from *telecommunications providers* to *communications providers* and (ii) to include intrastate communications. The base is thus expanded beyond circuit-switched voice and voice only applications, in a manner that is consistent with the distribution base. Revenues are derived from the same infrastructure that is funded through the program. This approach also implements technology neutrality on the contribution side, and does not discriminate between various methods of providing voice, or between voice and broadband. In particular it includes the infrastructure portion of VoIP, broadband, and wireless communications.

The layered approach taken here differs markedly from those taken in Congressional bills. Both the Stevens and Boucher/Terry bills require contributions from communications service providers. The Stevens bill defines these as providers of telecommunications service, broadband service, or VoIP, and hence keeps the focus on the application rather than the infrastructure. The Boucher/Terry bill defines communications service providers as voice providers who use telephone numbers or IP addresses (thus including VoIP) and providers of physical transmissions facilities that allows access to a network used for electronic communications; this definition is a hybrid between application and infrastructure.

The Stevens bill allows contributions to the Federal USF to be based on communications service revenue, numbers, and/or capacity. The Boucher/Terry bill allows contributions to the Federal USF to be based on communications service revenue and/or numbers, but does not define which services qualify as communications services when a communications service provider is defined by services other than voice. In contrast, this proposal maintains revenue as the basis for assessments, and defines communication services in a layered fashion. The numbers approach creates an artificial incentive to reduce the number of network identifiers, and network capacity is a poor proxy for the value of communications service provided.

Both bills allow the Federal USF to assess intrastate communications services (in addition to interstate and international), similar to this proposal. These differences are displayed in figure 2.

6.3. *Federal and State funds*

Finally, turn to the relationship between contributions to the Federal and State universal service funds. In 2006, twenty-two states had high-cost state USF funds. All such states assess intrastate revenues of ILECs, competitive LECs, and interexchange carriers; most also assess intrastate revenues of wireless providers, and a few assess intrastate revenues of some VoIP providers (The National Regulatory Research Institute, 2006). The states, however, have widely varying manners of distributing high-cost state USF support.

If the Federal USF is changed to assess intrastate services, then the National Association of Regulatory Utility Commissioners (NARUC), representing state public service commissioners, supports allowing states assess interstate and international services (National Association of Regulatory Utility Commissioners (NARUC), 2006). In addition, if the Federal USF is changed to assess broadband, NARUC would prefer that states have the ability to assess broadband.

Since the proposal in this paper allows the Federal USF to assess all communication services, a symmetric approach would be to allow State USFs to similarly assess all communication services. This can be

accomplished using this statute language:

> **(b)** *STATE USF CONTRIBUTIONS.- Section 254(f) of the Communications Act of 1934 (47 U.S.F. 254(f)) is amended-*
>> **(1)** *by striking "telecommunications carrier that provides interstate telecommunications" and inserting "provider of intrastate, interstate, or international communications"; and*
>> **(2)** *by adding at the end "Nothing in this subsection precludes a State from requiring contributions with respect to communications services for which contributions are required under subsection (d) if the primary place of use of which is within the State, regardless of where the services originate or terminate or through which the services transit.".*

First, this language changes the USF contribution base from *telecommunications* providers to *communications* providers. Secondly, it allows states to assess interstate and international communications services. To ensure that communication services can not be assessed by more than one State, services can only be assessed by the State of the primary place of use (the customer's residence or business address), an approach that was used in the *Mobile Telecommunications Sourcing Act* (U.S. Congress, 2000). Parity will thus exist between the Federal and State USF funds. The broadening of the Federal contribution base will decrease the assessment rate but include intrastate communications. If a State chooses to do so, it can similarly decrease its assessment rate but include interstate and international communications.

Finally, a corresponding change is required to one of the USF principles:

SEC. 7. EXPANSION OF UNIVERSAL SERVICE PRINCIPLES.

Section 254(b)(4) of the Communications Act of 1934 (47 U.S.C. 254(b)(4)) is amended to read as follows:

> *"(4) EQUITABLE AND NONDISCRIMINATORY CONTRIBUTIONS.- All providers of communications services should make an equitable and nondiscriminatory contribution to the preservation and advancement of universal service based on revenue derived from communications services without regard to whether such services are intrastate, interstate, or international in character.".*

Again, *telecommunication services* has been changed to *communication services*, and the restriction to interstate and international revenue has been removed.

This approach to state funds is similar to that taken in Congressional bills. Both the Stevens and Terry/Boucher bills allow the Federal USF to assess interstate and international communications services (in addition to intrastate), similar to the approach taken here. The Terry/Boucher bill allows state to assess the same base as the Federal fund, whereas the Stevens bill only allows states to assess communications and VoIP but not broadband. These differences are displayed in figure 2.

7. Conclusion

This paper has proposed a layered approach to restructuring the high-cost funds of the USF to reposition the Fund away from a focus on telephone service and toward an everything-over-IP world. Key elements of the proposal include a new definition of *communication services* and *communications infrastructure* to replace reliance on the out-dated definition of *telecommunication services*; restructuring of high-cost USF distributions from voice network infrastructure to communications infrastructure; service area reform to allow wireless and cable providers to use their own service areas and to remove incentives for cream-skimming; introduction of cost-saving measures; and revision of contributions methodology to allow both Federal and State USF to assess all communications services. A summary of this proposal, in comparison to current law and Congressional bills, is presented in figure 2. In order to complete any such approach, further consideration should be given to a cap on the high-cost funds, to finer disaggregation, and to the other portions of the Fund such as Lifeline and E-rate.

The principal idea in this proposal is the delineation of infrastructure from applications. Such a delineation is potentially of use more broadly within telecommunications policy. Wherever it has been of value to consider the distinction between telecommunication services from information services, this new definition of communication services may help make a case for substantially different regulation of infrastructure than of applications.

References

Abernathy, K. Q., 2005. Preserving universal service in the age of IP. Journal on Telecommunications & High Technology Law 3, 409.

Downes, T., Greenstein, S., 2006. Understanding why universal service obligations may be unnecessary: The private development of local Internet access markets. Discussion Papers Series 0615, Department of Economics, Tufts University, available at http://ideas.repec.org/p/tuf/tuftec/0615.html.

FCC, 1971. First Computer Inquiry, Final Decision, 28 FCC2d 267.

FCC, 1980. Second Computer Inquiry, Final Decision, 77 FCC2d 384.

FCC, 2002. FCC 02-77, Cable Modem Service Order. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-02-77A1.pdf.

FCC, 2005a. FCC 05-150, Internet over Wireline Facility Order. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-05-150A1.pdf.

FCC, 2005b. FCC 05-46, ETC Order. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-05-46A1.pdf.

FCC, 2007. DOC-279226, 2007 Universal Service Monitoring Report. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/DOC-279226A1.pdf.

FCC, 2008a. FCC 08-122, USF Cap Order. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08-122A1.pdf.

FCC, 2008b. FCC 08-4, Identical Support NPRM. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08-4A1.pdf.

FCC, 2008c. FCC 08-5, Reverse Auction NPRM. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-08-5A1.pdf.

Federal-State Joint Board on Universal Service, 2007. FCC 07J-4, USF Reform Recommended Decision. Available at http://hraunfoss.fcc.gov/edocs_public/attachmatch/FCC-07J-4A1.pdf.

Goldfarb, C. B., 2006. RL33034, Telecommunications Act: Competition, Innovation, and Reform. Tech. rep., Congressional Research Service.

Jayakar, K., Sawhney, H., 2007. Universal access in the information economy: Tracking policy innovations abroad. Tech. rep., Benton Foundation.

Jordan, S., 2007. A layered network approach to net neutrality. International Journal of Communication 1, 427–460.

Kruger, L. G., Gilroy, A. A., 2007. RL30719 Broadband Internet Access and the Digital Divide: Federal Assistance Programs. Tech. rep., Congressional Research Service.

Lessig, L., 2001. The Internet under siege. Foreign Policy 127, 56–65.

Marcus, J., 2005. Universal service in a changing world - regulatory and policy issues. IEEE Communications Magazine 43 (1), 16–17.

National Association of Regulatory Utility Commissioners (NARUC), 2006. Testimony of the Honorable Philip Jones, Commissioner, Washington Utilities and Transportation Commission before the Committee on Commerce, Science and Transportation of the United States Senate.

Oh, K. J., 2006. Completing the connection: Achieving universal service through municipal Wi-Fi. Duke Law & Technology Review 1.

Prieger, J. E., 2003. The supply side of the digital divide: Is there equal availability in the broadband Internet access market? Economic Inquiry 41, 346–363.

Schejter, A. M., 2007. From all my teachers I have grown wise, and from my students more than anyone else: What lessons can the U.S. learn from broadband policies in Europe? In: Proceedings of the Telecommunication Policy Research Conference.

Sicker, D., 2003. Applying a layered policy model to IP based voice services. In: Proceedings of the 36th Annual Hawaii International Conference on System Sciences. pp. 151–160.

Solum, L. B., Chung, M., 2003. The layers principle: Internet architecture and the law. U San Diego Public Law Research Paper No. 55Available at http://ssrn.com/abstract=416263.

Taylor, R. D., 2007. Time for change: Transforming funding for broadband universal service. Tech. rep.,

Benton Universal Service Project.

The National Regulatory Research Institute, 2006. State universal service funding mechanisms: results of the NRRI's 2005-2006 survey.

Turner, S. D., 2006. Universal service reform & convergence: USF policy for the 21st century. In: Proceedings of the Telecommunications Policy Research Conference (TPRC).

U.S. Congress, 2000. H.R.4391.PP, Mobile Telephone Sourcing Act, 106th Congress, Public Law 106-252. Available at http://thomas.loc.gov/cgi-bin/bdquery/z?d106:h.r.04391:.

U.S. Congress, 2004. S.1380.RS, Rural Universal Service Equity Act of 2003, 108th Congress. Available at http://thomas.loc.gov/cgi-bin/bdquery/z?d108:s.001380:.

U.S. Congress, 2006. H.R.5252.EH, Advanced Telecommunications and Opportunities Reform Act, 109th Congress. Available at http://thomas.loc.gov/cgi-bin/bdquery/z?d109:HR05252:.

U.S. Congress, 2007a. H.R.2054.IS, Universal Service Reform Act of 2007, 110th Congress. Available at http://thomas.loc.gov/cgi-bin/bdquery/z?d110:h.r.02054:.

U.S. Congress, 2007b. S.101.IS, Universal Service for Americans Act, 110th Congress. Available at http://thomas.loc.gov/cgi-bin/bdquery/z?d110:S.00101:.

U.S. General Accounting Office, 2002. GAO-02-187, Federal and State Universal Service Programs and Challenges to Funding.

Weinberg, J., 2001. Internet telephony regulation. In: McKnight, L. W., Lehr, W., Clark, D. D. (Eds.), Internet Telephony. MIT Press.

Werbach, K. D., 2002. A layered model for Internet policy. Journal on Telecommunications and High-Tech Law 1, available at http://ssrn.com/abstract=648581.

Whitt, R. S., 2004. A horizontal leap forward: Formulating a new communications public policy framework based on the network layers model. Federal communications law journal 56, 587–672.